

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 15:03:13 ; Search time 3307 Seconds  
(without alignments)  
9847.608 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119  
Sequence: 1 atgataaccagagctccat.....ttcagcaggttgaataaa 1119

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_cm.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_un.\*  
28: em\_vl.\*  
29: em\_hg\_hum.\*  
30: em\_hg\_inv.\*  
31: em\_hg\_inv.\*  
32: em\_hg\_inv.\*  
33: em\_hg\_mus.\*  
34: em\_hg\_pln.\*  
35: em\_hg\_rnd.\*  
36: em\_hg\_rnd.\*  
37: em\_hg\_vrt.\*  
38: em\_sy.\*  
39: em\_hg\_hum.\*  
40: em\_hg\_mus.\*  
41: em\_hg\_inv.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	6 AX036302	AX036302 Sequence
2	1119	100.0	1119	6 AX038912	AX038912 Sequence
3	1119	100.0	1119	6 AX050487	AX050487 Sequence
4	1119	100.0	1119	6 AX039343	AX039343 Sequence
5	1119	100.0	1697	1 EC6GPE	X64451 E.coli gcpE
6	1119	100.0	13176	1 AE000338	AE000338 Escherich
7	1093.4	97.7	11521	1 AE003515	AE003515 Escherich
8	1093.4	97.7	11521	1 AE005481	AE005481 Escherich
9	1093.4	97.7	296827	1 AR027993	AR027993 Sequence
10	1093.4	90.3	1010	6 AR208328	AR208328 Sequence
11	1010	90.3	1010	6 AR208328	AR208328 Sequence
12	887	79.3	23647	1 AE008814	AE008814 Salmonell
13	885.4	79.1	145050	1 AL627275	AL627275 Salmonell
14	710	63.4	4775	1 D90881	D90881 E.coli geno
15	690	61.7	16214	1 D90880	D90880 E.coli geno
16	685.4	59.4	11824	1 AE004161	AE004161 Mus muscu
17	665	56.4	1345	1 PS067931	PS067931 Vibrioc
18	631.6	56.4	1345	6 AR208326	AR208326 Sequence
19	631.6	56.4	1345	6 AR208326	AR208326 Sequence
20	631.6	56.4	1345	6 AR208326	AR208326 Sequence
21	622	55.6	10507	1 U32721	U32721 Haemophilus
22	620.6	55.5	11161	1 AE013738	AE013738 Yersinia
23	620.6	55.5	204050	1 AJ414154	AJ414154 Yersinia
24	619.8	55.4	1079	6 AR027992	AR027992 Sequence
25	619.8	55.4	1079	6 AR027992	AR027992 Sequence
26	603.6	53.9	11268	1 AE006237	AE006237 Pasteurel
27	534.6	47.8	886	6 AR027995	AR027995 Sequence
28	534.6	47.8	886	6 AR027995	AR027995 Sequence
29	515.4	46.1	17148	1 AE004798	AE004798 Pseudomon
30	403.6	36.1	347550	1 AP001118	AP001118 Buchnera
31	388.8	34.7	9608	1 AE014104	AE014104 Caulobact
32	383.6	34.3	9999	1 AE005761	AE005761 Homo sapi
33	382.6	34.2	33140	6 AX067448	AX067448 Homo sapi
34	317.2	28.3	987	6 HSA335686	AJ332686 Homo sapi
35	293.4	26.2	1083	6 AX433277	D29116 Bacillus su
36	277.2	24.8	218470	1 BSUB0013	D84432 Bacillus su
37	277.2	24.8	282700	1 BAC01642	AR027994 Sequence
38	269	24.0	886	6 AR027994	AR027994 Sequence
39	269	23.6	43632	1 SC6A5	AL048485 Streptomy
40	263.6	23.1	39744	1 SC5H4	AL355913 Streptomy
41	258.2	22.8	6157	6 AX416761	AX416761 Sequence
42	254.6	22.5	1107	6 AX413788	AX413788 Sequence
43	251.4	22.5	1107	6 AX415815	AX415815 Sequence
44	251.4	22.5	240050	1 AL591979	AL591979 Listeria
45	251.4	22.5	240050	1 AL591979	AL591979 Listeria

ALIGNMENTS

RESULT 1  
LOCUS AX036302  
DEFINITION Sequence 29 from Patent EP1043403.  
ACCESSION AX036302  
VERSION AX036302.1 GI:11225912  
KEYWORDS  
SOURCE Escherichia coli.  
ORGANISM Escherichia coli.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
FEATURES  
source 1..1119  
/organism="Escherichia coli"  
/db\_xref="taxon:562"

BASE COUNT 272 a 279 c 317 g 251 t

Query Match 100.0%; Score 1119; DB 6; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 9.9e-271;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGCATACCAAGGCTCCATTCATCAAGCTAGAAAAATCAACAGCTATTTACGTTGGGAATGTG 60  
1 ATGCATACCAAGGCTCCATTCATCAAGCTAGAAAAATCAACAGCTATTTACGTTGGGAATGTG 60  
61 CCGATTGGCGATGGTGGCTGATCGGCTGACATGACATCAATACGCGTACAGACAGAC 120  
61 CCGATTGGCGATGGTGGCTGATCGGCTGACATGACATCAATACGCGTACAGACAGAC 120  
121 GTGGAACCAAGCTCAATGAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGCTG 180  
121 GTGGAACCAAGCTCAATGAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGCTG 180  
181 GTATCCGTACCGGATGAGGAGGCGGAGAGGCTTCAACATCATCAACAGCAGGTTAAC 240  
181 GTATCCGTACCGGATGAGGAGGCGGAGAGGCTTCAACATCATCAACAGCAGGTTAAC 240  
241 GTGCGCTGCTGGCTGACATTCATCTGATTCGATTCGCTGAAAGTACCGGAATAC 300  
241 GTGCGCTGCTGGCTGACATTCATCTGATTCGATTCGCTGAAAGTACCGGAATAC 300  
301 GCGTCGATTTGCTGCTGATTAACCTGCAATATCGGTAATGAAGACGCTATTCGATG 360  
301 GCGTCGATTTGCTGCTGATTAACCTGCAATATCGGTAATGAAGACGCTATTCGATG 360  
361 GTGGTTGACTGTGCGCGGATAAAAACATTCGATTCGATTCGCTGAAAGTACCGGAATG 420  
361 GTGGTTGACTGTGCGCGGATAAAAACATTCGATTCGATTCGCTGAAAGTACCGGAATG 420  
421 CTGGAAGAAAGATCTCAAGAAAGTANGGCGACCGGCGGAGCGCTTGGCTGGAATCT 480  
421 CTGGAAGAAAGATCTCAAGAAAGTANGGCGACCGGCGGAGCGCTTGGCTGGAATCT 480  
481 GCCATGCTCATGTTGATGATCTGATGCGCTGATGCTGATGCTGATGCTGATGCTG 540  
481 GCCATGCTCATGTTGATGATCTGATGCGCTGATGCTGATGCTGATGCTGATGCTG 540  
541 AAAGCTGTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
541 AAAGCTGTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
601 CAGCGCTTGCATCTGGGAGAGACCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
601 CAGCGCTTGCATCTGGGAGAGACCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
661 GCCATTGCTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
661 GCCATTGCTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
721 GCGGCGATCCGCTGCAAGAGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
721 GCGGCGATCCGCTGCAAGAGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
781 CATTGCGAGAGGATCAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
781 CATTGCGAGAGGATCAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
841 ATGCGTACGCTTAAAGCGCTGAGAGAGCGCTGAGAGATATCATCTCCGATGAGCTT 900  
841 ATGCGTACGCTTAAAGCGCTGAGAGAGCGCTGAGAGATATCATCTCCGATGAGCTT 900  
901 TCGATTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
901 TCGATTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
961 ACCGCGGCAACAGAAAAAGGCGCTTATGAAGATGCGGCGCAAGACCGCTGCTGAC 1020  
961 ACCGCGGCAACAGAAAAAGGCGCTTATGAAGATGCGGCGCAAGACCGCTGCTGAC 1020

Db 961 ACCGCGGCAACAGAAAAAGGCGCTTATGAAGATGCGGCGCAAGACCGCTGCTGAC 1020  
Qy 1021 AACACGATATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1080  
Db 1021 AACACGATATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1080  
Qy 1081 GAAGCGCTGCAATTTGACGCTTACGAGGTTGAAAAATTA 1119  
Db 1081 GAAGCGCTGCAATTTGACGCTTACGAGGTTGAAAAATTA 1119

RESULT 2  
AX038912 1119 bp DNA linear PAT 16-NOV-2000  
LOCUS AX038912  
DEFINITION Sequence 29 from Patent WO0061793.  
ACCESSION AX038912  
VERSION AX038912.1 GI:11228221  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.

REFERENCE  
1 (bases 1 to 1119)  
AUTHORS Loferer, H. and Jacob, A.  
TITLE Novel method for identifying antibacterial compounds  
JOURNAL Patent: WO 0061793-A 29-19-OCT-2000;  
GPC BIOTECH AG (DE); LOFERER HANNES (DE); JACOBI ALEXANDER (DE)

FEATURES  
source  
1..1119  
/organism="Escherichia coli"  
/db\_xref="taxon:562"

BASE COUNT 272 a 279 c 317 g 251 t

ORIGIN

Query Match 100.0%; Score 1119; DB 6; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 9.9e-271;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGCATACCAAGGCTCCATTCATCAAGCTAGAAAAATCAACAGCTATTTACGTTGGGAATGTG 60  
1 ATGCATACCAAGGCTCCATTCATCAAGCTAGAAAAATCAACAGCTATTTACGTTGGGAATGTG 60  
61 CCGATTGGCGATGGTGGCTGATCGGCTGACATGACATCAATACGCGTACAGACAGAC 120  
61 CCGATTGGCGATGGTGGCTGATCGGCTGACATGACATCAATACGCGTACAGACAGAC 120  
121 GTGGAAGCAAGGCTCAATCAATCAAGGCGCTGGAACGCGCTTGGCGCTGATATCGCTG 180  
121 GTGGAAGCAAGGCTCAATCAATCAAGGCGCTGGAACGCGCTTGGCGCTGATATCGCTG 180  
181 GTATCCGTACCGGATGAGGAGGCGGAGAGGCTTCAACATCATCAACAGCAGGTTAAC 240  
181 GTATCCGTACCGGATGAGGAGGCGGAGAGGCTTCAACATCATCAACAGCAGGTTAAC 240  
241 GTGCGCTGCTGGCTGACATTCATCTGATTCGATTCGCTGAAAGTACCGGAATAC 300  
241 GTGCGCTGCTGGCTGACATTCATCTGATTCGATTCGCTGAAAGTACCGGAATAC 300  
301 GCGTCGATTTGCTGCTGATTAACCTGCAATATCGGTAATGAAGACGCTATTCGATG 360  
301 GCGTCGATTTGCTGCTGATTAACCTGCAATATCGGTAATGAAGACGCTATTCGATG 360  
361 GTGGTTGACTGTGCGCGGATAAAAACATTCGATTCGATTCGCTGAAAGTACCGGAATG 420  
361 GTGGTTGACTGTGCGCGGATAAAAACATTCGATTCGATTCGCTGAAAGTACCGGAATG 420  
421 CTGGAAGAAAGATCTCAAGAAAGTANGGCGACCGGCGGAGCGCTTGGCTGGAATCT 480  
421 CTGGAAGAAAGATCTCAAGAAAGTANGGCGACCGGCGGAGCGCTTGGCTGGAATCT 480  
481 GCCATGCTCATGTTGATGATCTGATGCGCTGATGCTGATGCTGATGCTGATGCTG 540  
481 GCCATGCTCATGTTGATGATCTGATGCGCTGATGCTGATGCTGATGCTGATGCTG 540

QY 541 AAAGCGTGTGACGCTCTCCGCGGTGAGCTTATGCTTGTGGCAAAACAGATGAT 600  
 DB 541 AAAGCGTGTGACGCTCTCCGCGGTGAGCTTATGCTTGTGGCAAAACAGATGAT 600  
 QY 601 CAGCGGTGTGATCTGGGGATCACCGAGCGGTGTGGCGGCGGAGGAGTAAATCC 660  
 DB 601 CAGCGGTGTGATCTGGGGATCACCGAGCGGTGTGGCGGCGGAGGAGTAAATCC 660  
 QY 661 GCCATTGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGGCGGTATCCGTG 720  
 DB 661 GCCATTGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGGCGGTATCCGTG 720  
 QY 721 CGGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATGCTGCGATC 780  
 DB 721 CGGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATGCTGCGATC 780  
 QY 781 CGTTGCGAGGATCACTTCACTGCGCTGCGGACCTGTTGCGCTGAGAAATTTGATGTT 840  
 DB 781 CGTTGCGAGGATCACTTCACTGCGCTGCGGACCTGTTGCGCTGAGAAATTTGATGTT 840  
 QY 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTGGAATATCATCTCCGATGGAGCTT 900  
 DB 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTGGAATATCATCTCCGATGGAGCTT 900  
 QY 901 TCGATTATCGGCTGGGTGATGAGTGGCCAGGTGAGGCGGTGTTTACACTGCGGCTC 960  
 DB 901 TCGATTATCGGCTGGGTGATGAGTGGCCAGGTGAGGCGGTGTTTACACTGCGGCTC 960  
 QY 961 ACCGCGGCAACAGAAAGCGGCTCTATGAGATGGCGTGGCGAAAGACCTGCTGGAC 1020  
 DB 961 ACCGCGGCAACAGAAAGCGGCTCTATGAGATGGCGTGGCGAAAGACCTGCTGGAC 1020  
 QY 1021 AACCAAGATATGATGACCACTGGAAGCAGCATCTGCGAAACGACGATGGAC 1080  
 DB 1021 AACCAAGATATGATGACCACTGGAAGCAGCATCTGCGAAACGACGATGGAC 1080  
 QY 1081 GAAGCGCTGCAATGAGCTTCAGAGGTTGAAAAATAA 1119  
 DB 1081 GAAGCGCTGCAATGAGCTTCAGAGGTTGAAAAATAA 1119  
 RESULT 3  
 AX050487 1119 bp DNA linear PAT 12-JAN-2001  
 LOCUS Sequence 1 from Patent WO0072022.  
 ACCESSION AX050487  
 VERSION AX050487.1 GI:12226691  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 1 (bases 1 to 1119)  
 JOURNAL  
 Jomaa, H.  
 Use of genes of the deoxy-d-xylose phosphate biosynthetic pathway  
 for altering the concentration of isoprenoid  
 Patent: WO 0072022-A 1 30-NOV-2000;  
 Jomaa, Hassan (DE)  
 FEATURES  
 source  
 1..1119  
 Location/Qualifiers  
 /organism="Escherichia coli"  
 /db\_xref="taxon:562"  
 1..1119  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC21801.1"  
 /db\_xref="gi:12226692"  
 /translation="MHNOPIORRSTRIVYGNPISGAPLIVASMTNFTTVEAT  
 VNOTKALEKRGADIVRVSVPTMDAAEFKLKQOVNPLVDIHEDYIAIKVAYGV  
 DCLRTNPNIGNERIKRVYVCDARKNRPIRIGVAGSLDLEKYEGRPPQALIES  
 AMRHVDHLDRLNFDQFKVSKASDVFLAVESTRLAKOIIDPLHLGITIEGANSNAV

KSAIGLLISBGIDTLRLVSLAADPEEIKVGFLLKSLRISRGINFACPTCSRO  
 EFDVISTVALAEQREDIITPMDSVSIIGVNGPGLALVSTLGTGNKRSGLYEDGV  
 RKRLNNMMDLEARIKAKASQLEARRIDVQYVER"  
 BASE COUNT 272 a 279 c 317 g 251 t  
 ORIGIN  
 Query Match 100.0%; Score 1119; DB 6; Length 1119;  
 Best Local Similarity 100.0%; Pred. No. 9,9e-271;  
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGCATACCAAGGCTCCAAATTCAGAGTAAGAAATCAACAGATATTTCAGTTGGAAATCTG 60  
 DB 1 ATGCATACCAAGGCTCCAAATTCAGAGTAAGAAATCAACAGATATTTCAGTTGGAAATCTG 60  
 QY 61 CCGATTGGCATAGGTGCTCCCATGCGGTACAGTCATGACCATATAGCGGTAGCAGAC 120  
 DB 61 CCGATTGGCATAGGTGCTCCCATGCGGTACAGTCATGACCATATAGCGGTAGCAGAC 120  
 QY 121 GTGCAAGCAACGCTCAATCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATATCGTCGT 180  
 DB 121 GTGCAAGCAACGCTCAATCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATATCGTCGT 180  
 QY 181 GTATCCGTACCGACAGATGAGCGCGGAGAGCGTTCAACATCATCAACAGCAGATTAC 240  
 DB 181 GTATCCGTACCGACAGATGAGCGCGGAGAGCGTTCAACATCATCAACAGCAGATTAC 240  
 QY 241 GTGCGCGGTGGGTGACATCACTGCTGATGCTGATGCTGAAAGTATGCGGAATAC 300  
 DB 241 GTGCGCGGTGGGTGACATCACTGCTGATGCTGATGCTGAAAGTATGCGGAATAC 300  
 QY 301 GCGCTCATGTTGCTGCTATTAACTTGGCAATATCGGTATGAAGACGTTATTCGATG 360  
 DB 301 GCGCTCATGTTGCTGCTATTAACTTGGCAATATCGGTATGAAGACGTTATTCGATG 360  
 QY 361 GTGTTGACTGTGGCGGAGTAAACATTCGCAATCCGTTATGGCGTTAATGCGCGATCG 420  
 DB 361 GTGTTGACTGTGGCGGAGTAAACATTCGCAATCCGTTATGGCGTTAATGCGCGATCG 420  
 QY 421 CTGCAAAAGATCTGCAAGAAAGTATGGCGAGCGAGCGCGGCTTCTGGAATCT 480  
 DB 421 CTGCAAAAGATCTGCAAGAAAGTATGGCGAGCGAGCGCGGCTTCTGGAATCT 480  
 QY 481 GCCATGCTCATGTTGATCATCTGATCGCCTGAACCTTCATCACTTAAAGTCAAGCTG 540  
 DB 481 GCCATGCTCATGTTGATCATCTGATCGCCTGAACCTTCATCACTTAAAGTCAAGCTG 540  
 QY 541 AAAGCGCTGACGCTCTCTGCTGCTGCTGATGCTTATGCTTGGCGCAAAACAGATCGAT 600  
 DB 541 AAAGCGCTGACGCTCTCTGCTGCTGCTGATGCTTATGCTTGGCGCAAAACAGATCGAT 600  
 QY 601 CAGCGGTGTGATCTGGGGATCACCGAGCGGTGTGGCGGCGGAGGAGTAAATCC 660  
 DB 601 CAGCGGTGTGATCTGGGGATCACCGAGCGGTGTGGCGGCGGAGGAGTAAATCC 660  
 QY 661 GCCATTGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGGCGGTATCCGTG 720  
 DB 661 GCCATTGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGGCGGTATCCGTG 720  
 QY 721 CGGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATGCTGCGATC 780  
 DB 721 CGGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATGCTGCGATC 780  
 QY 781 CGTTGCGAGGATCACTTCACTGCGCTGCGGACCTGTTGCGCTGAGAAATTTGATGTT 840  
 DB 781 CGTTGCGAGGATCACTTCACTGCGCTGCGGACCTGTTGCGCTGAGAAATTTGATGTT 840  
 QY 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTGGAATATCATCTCCGATGGAGCTT 900  
 DB 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTGGAATATCATCTCCGATGGAGCTT 900  
 QY 901 TCGATTATCGGCTGGGTGATGAGTGGCCAGGTGAGGCGGTGTTTACACTGCGGCTC 960  
 DB 901 TCGATTATCGGCTGGGTGATGAGTGGCCAGGTGAGGCGGTGTTTACACTGCGGCTC 960





JOURNAL Submitted (17-FEB-1992) J. Parker, Southern Illinois Univ, Dept of Microbiology, S 1 U, Carbondale IL 62901, USA

REFERENCE 2 (bases 1 to 1697)  
Baker, J., Franklin, D. B. and Parker, J.  
Sequence and characterization of the *gcpE* gene of *Escherichia coli*

TITLE *gcpE* Microbiol. Lett. 94, 175-180 (1992)

JOURNAL See also J01629 & M11843.

COMMENT Location/Qualifiers

FEATURES

source

1. 1697  
/organism="Escherichia coli"  
/strain="K12"  
/db\_xref="taxon:562"  
/sub\_clone="PSEA11"  
/cell\_line="JF298"  
/clone\_1lb="lambda transducing phage dguab-2"  
/note="ORF"  
/codon\_start=2  
/transl\_table=1  
/protein\_id="CAA45782.1"  
/db\_xref="GI:41541"  
/db\_xref="SWISS-PROT:P27434"  
/translation="TTTDPATSTPPASVDTTANTPTATPAPAPAVDPQNAVSP.  
SQAVDTATAPATPAPATPDPGAPLPDQGVTPPADPALVNFATDCMLEVTAT  
GKLFSGMQRKDNLTGAPYKLTIGAPAAVYIQYQKRPVLSRTIRNQVARIIL  
NAEOSPAQ"  
535..1653  
/gene="gcpE"  
535..1653  
/gene="gcpE"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAA45783.1"  
/db\_xref="GI:41542"  
/db\_xref="SWISS-PROT:P27433"  
/translation="MHNQPIORRSTRIYGNVPIGDAPIVASMTNRTTDEAT  
VNOIKLEAVGADIVRYSPPTMDAAEFLIKQVNVPLVDIHEFVIAIKVAGY  
DCRLNPGINBERIRNRYVDCARDKNPTRIGVAGSLEKLOEYEPPOALLS  
AMRHVDHDLNFDOKFVSKASDVPLAVESRYLLAKIIDPLHIGITEAGASGAV  
KSAIIGLLSLSEIGDTLVSLIADVVEIKYGPDLIKSLIRSRGINFIACPCSRQ  
EFVIGTVNALEQRLEDITTPMDVSLIGCVVGPBALYSTIGVTKGKKSGVLEDGY  
RKDLNNDIMIDLEARIKASOLDEARIDVOVER"

BASE COUNT 417 a 446 c 469 g 365 t

Query Match 100.0%; Score 1119; DB 1; Length 1697;  
Best Local Similarity 100.0%; Pred. No. 1e-270; Indels 0; Gaps 0;  
Matches 1119; Conservative 0; Mismatches 0;

OY 1 ATGCTATACGAGGCTCCCAATTCACAGTAGAATAACACAGCTATTTACGTTGGATGTG 60  
DB 535 ATGCTATACGAGGCTCCCAATTCACAGTAGAATAACACAGCTATTTACGTTGGATGTG 594  
OY 61 CGGATTGGGATGTGCTCCATGCGCGGTACAGTCATGACCAATACGCGGTACAGACAG 120  
DB 595 CGGATTGGGATGTGCTCCATGCGCGGTACAGTCATGACCAATACGCGGTACAGACAG 654  
OY 121 GTTCGAAGCAAGGTCATCAATCAAGGCGGTGAAGCGGTTGGCGCTGATATGTCGCT 180  
DB 655 GTTCGAAGCAAGGTCATCAATCAAGGCGGTGAAGCGGTTGGCGCTGATATGTCGCT 714  
OY 181 GTATCCGTACGACGATGACGCGGAGAAAGCGTTCAAACTCATCAACACAGAGGTAAAC 240  
DB 715 GTATCCGTACGACGATGACGCGGAGAAAGCGTTCAAACTCATCAACACAGAGGTAAAC 774  
OY 241 GTCCCGCTGTGCTGATCACTTCGACTATGCGATTGGCTGAAGTACGCGAATAC 300  
DB 775 GTCCCGCTGTGCTGATCACTTCGACTATGCGATTGGCTGAAGTACGCGAATAC 834  
OY 301 GGGGTGATGTGCTGCTATTAACCTGGCAATATGCTGATGAAGACGCTATTGGCATG 360  
DB 835 GGGGTGATGTGCTGCTATTAACCTGGCAATATGCTGATGAAGACGCTATTGGCATG 894

OY 361 GTGCTGACTGTGGCGCGATTAACCAATTCGATCCGATATTTGGCTTAACCGCGGATG 420  
DB 895 GTGCTGACTGTGGCGCGATTAACCAATTCGATCCGATATTTGGCTTAACCGCGGATG 954  
OY 421 CTGAAAAAAGATCTGCAAGAAAGATATGCGAACCGACGCGCGAGCGTGTGATCT 480  
DB 955 CTGAAAAAAGATCTGCAAGAAAGATATGCGAACCGACGCGCGAGCGTGTGATCT 1014  
OY 481 GCCATGCTCATGTTGATCATCTGATCCGCTGAATCTGATCATGTTCAAAAGTACGCTG 540  
DB 1015 GCCATGCTCATGTTGATCATCTGATCCGCTGAATCTGATCATGTTCAAAAGTACGCTG 1074  
OY 541 AAAGCTGTGACGCTCTTCTCTGCTGCTGATGATCTTATGCTTTGCTGCGAAAGCATGAT 600  
DB 1075 AAAGCTGTGACGCTCTTCTCTGCTGCTGATGATCTTATGCTTTGCTGCGAAAGCATGAT 1134  
OY 601 CAGCGCTGCTATCTGGGATCACCGAAGCGGTGCGCGACGCGGCGGTAAATTC 660  
DB 1135 CAGCGCTGCTATCTGGGATCACCGAAGCGGTGCGCGACGCGGCGGTAAATTC 1194  
OY 661 GCGATGTTAGTCTGCTGCTGCTGCTGAAGGATCGCGACAGCTGCGGTATCGCTG 720  
DB 1195 GCGATGTTAGTCTGCTGCTGCTGCTGCTGAAGGATCGCGACAGCTGCGGTATCGCTG 1254  
OY 721 GCGCGCATTCGCTGCAAGAGATCAAGCGGTTGATATTTGAATCGCTGCTGATC 780  
DB 1255 GCGCGCATTCGCTGCAAGAGATCAAGCGGTTGATATTTGAATCGCTGCTGATC 1314  
OY 781 CGTTGCGAGAGATCAAGTTCATCGCTGCGCGACCTGTTGCGCTGAGATTTATGTT 840  
DB 1315 CGTTGCGAGAGATCAAGTTCATCGCTGCGCGACCTGTTGCGCTGAGATTTATGTT 1374  
OY 841 ATCGGTACGCTTAACGCGCTGAGCAACCGCTGGAAGATATCATCTCCGATGACGTT 900  
DB 1375 ATCGGTACGCTTAACGCGCTGAGCAACCGCTGGAAGATATCATCTCCGATGACGTT 1434  
OY 901 TCGATTTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
DB 1435 TCGATTTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1494  
OY 961 ACCGCGGCAACAGAAAGAGCGGCTTATGAAGATGCGGTGGCAAGACGCTGAGAC 1020  
DB 1495 ACCGCGGCAACAGAAAGAGCGGCTTATGAAGATGCGGTGGCAAGACGCTGAGAC 1554  
OY 1021 AACAAGCATATGATGACGACGCTGGAAGCAAGCATGCTGCGAAGGCACTGACGCTGAC 1080  
DB 1555 AACAAGCATATGATGACGACGCTGGAAGCAAGCATGCTGCGAAGGCACTGACGCTGAC 1614  
OY 1081 GAAGCGCTGCAATGACGCTGACGCTGCAAAATTA 1119  
DB 1615 GAAGCGCTGCAATGACGCTGACGCTGCAAAATTA 1653

RESULT 6  
AE000338/c 13176 bp DNA linear BCT 01-DEC-2000  
LOCUS Escherichia coli K12 MG1655 section 228 of 400 of the complete genome.  
DEFINITION  
ACCESSION AE000338 U00096  
VERSION AE000338.1 GI:1788862  
KEYWORDS  
SOURCE Escherichia coli K12.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 13176)  
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
The complete genome sequence of *Escherichia coli* K-12  
Science 277 (5331), 1453-1474 (1997)

JOURNAL MEDLINE 97426617

PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 9278503  
 2 (bases 1 to 13176)  
 Blattner, F.R.  
 Direct Submission  
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 3 (bases 1 to 13176)  
 Blattner, F.R.  
 Direct Submission  
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 4 (bases 1 to 13176)  
 Plunkett, G. III  
 Direct Submission  
 Submitted (13-06-1998) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 This sequence was determined by the E. coli Genome Project at the  
 University of Wisconsin-Madison (Frederick R. Blattner, director).  
 Projected by NIH grants HG00301 and HG01428 (from the Human Genome  
 Project and NCHGR). The entire sequence was independently  
 determined from E. coli K12 strain MG1655. Predicted open reading  
 frames were determined using Genemark software, kindly supplied by  
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,  
 30332 (e-mail: markborov@gatech.edu). Open reading frames that  
 have been correlated with genetic loci are being annotated with CG  
 Site Nos., unique ID nos. for the genes in the E. coli Genetic  
 Mary Berlyn. A public version of the database is accessible  
 (http://cgsc.biology.yale.edu). Annotation of the genome is an  
 ongoing task whose goal is to make the genome sequence more useful  
 by correlating it with other data. Comments to the authors are  
 appreciated. Updated information will be available at the E. coli  
 Genome Project's World Wide Web site  
 (http://www.genetics.wisc.edu). \*\* The E. coli K12 sequence and  
 its annotations are periodically updated; this is version M54. No  
 and products: all new functional assignments courtesy of Monica  
 Riley; added promoters, protein binding sites, and repeated  
 sequences described in reference 1. The unique numeric identifiers  
 beginning with a lowercase 'b' assigned to each gene (protein- or  
 RNA-encoding) are now designated as gene synonyms instead of  
 labels. This should allow them to be searched for in Entrez as gene  
 names.

FEATURES  
 source  
 location/qualifiers  
 1. 13176  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"  
 /db\_xref="taxon:83333"  
 complement(49..81)  
 /note="factor Sigma70; promoter hss; documented +1 at  
 2638662"  
 complement(86..1204)  
 /gene="gcpE"  
 /note="b2515"  
 complement(86..1204)  
 /gene="gcpE"  
 /function="orf: Unknown"  
 /note="b3072; 100 pct identical to GCP\_ECOLI SW: P27433"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AAC75568.1"  
 /db\_xref="GI:1788863"  
 /translation="MHQAPQRRKSTRIVYGNPJDGAPLAVQSMNTNRTTDEVT  
 VNOIKALEMGADIVRSVPMDAEAKFKLKQOVNPLVDIHFDYRLKAEVGS  
 DCIRNNGNENIRKRVNVCARDKNPIPIRGVNSGLDQEKGEPTPALES  
 AMRHVDLRINDFQFVSVKADSVFLAVESYRLAKQIDOPHLGITTEGCRSAV  
 KSAIGLILNSEGIGDILRVSLADPVEIKVGDILKSLIRSRGINFLACTCSRO

promoter  
 gene  
 CDS

gene  
 CDS  
 EDVIGTVALEQRLEDTITPMDSYIICGVNCGEALVSLCTVTSNKKSGLYEDGV  
 RKDLNDNDMDIQDEAETIRAKASQLEARRIDVOQVER"  
 complement(1231..2244)  
 /gene="yfga"  
 /note="b2516"  
 complement(1231..2244)  
 /gene="yfga"  
 /function="putative membrane; Not classified"  
 /note="f337; 100 pct identical to YFGA\_ECOLI SW: P27434"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane protein"  
 /protein\_id="AAC75569.1"  
 /db\_xref="GI:1788864"  
 /translation="MNEPAPDHQNEALTTGARLNRARQGLISQVAERLKYSTV  
 RQIEERKAPADLASTFRLYIRSVARLVAHPEELIPGLERQAPLRARVAPMSFL  
 GKRKRKRDGMWTFWLVLEVVIGSGAMWQDRAQOEITTMADQSSALSSNSQ  
 GDSVPLNTSTTDDPATSTPPASVPTTANTQTPAVTAPAVAPDPOQNAVYPSQANV  
 DTATPAPATTPDGAAPLPDQAGVTPVADPALVMTADQMLEVTPATSKLP  
 SGMQRKGNLNLGQAPYKIKIGAPAAVQIQGKRPVLSFRITNOVARLTLNAECS  
 PAQ"  
 complement(1288..1316)  
 /gene="yfga"  
 /note="factor Sigma70; predicted +1 start at 2639901"  
 complement(1360..1389)  
 /gene="yfga"  
 /note="factor Sigma70; predicted +1 start at 2639973"  
 complement(2529..3683)  
 /gene="yfgb"  
 /note="b2517"  
 complement(2529..3683)  
 /gene="yfgb"  
 /function="orf: Unknown"  
 /note="f384; 100 pct identical to YFGB\_ECOLI SW: P36979"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AAC75570.1"  
 /db\_xref="GI:1788865"  
 /translation="MSQQLVPEVNTTKDKINLMDLNQOMREFFKDGKPPRADO  
 VMKMYHYCDNDEMNDINKVLGKLEAEIRAEVEEORSSGTTKMAIYAGDO  
 RVEVTVIPEDRATLCVSOVGALEKPCSTAOGFNRLRYSEITIGVWRAATV  
 AAKVTGQRPITNVMMGEPPLNNVNPAMKYNIDSCFGLSKRRTVLSGVVPA  
 LDKRGMDIVLAISLHAPNDEIRDEIVPKRNINLETFLAVRRLTESNMQOQRT  
 IEVYMLDVNDGTGEHQAOLAEIKDKPCKINILPMPEPCAPYGRSNSNRIDRFKVL  
 MSVGETTIVKKTGSDIDACGQLAGVDIDRTKRLRKMGAEADIKAV"  
 complement(3778..3806)  
 /note="factor Sigma70; predicted +1 start at 2642391"  
 complement(3786..3814)  
 /note="factor Sigma70; predicted +1 start at 2642399"  
 complement(3833..4264)  
 /gene="ndk"  
 /note="b2518"  
 complement(3833..4264)  
 /gene="ndk"  
 /EC\_number="2.7.4.6"  
 /function="enzyme; Purine ribonucleotide biosynthesis"  
 /note="f143; 100 pct identical to NDK\_ECOLI SW: P24233"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nucleoside diphosphate kinase"  
 /protein\_id="AAC75571.1"  
 /db\_xref="GI:1788866"  
 /translation="MAIRTFESIIRPNVAKNVIGNIFANFLAAGFYVGTGKMLHTV  
 EQARFVAEHQDKPFDGLVEMTSGPIVSVGEENAVQRRRLDGLATNPANALAGT  
 IRADYADSLTEGTGSDSVSARRELAIFGEEVCPRT"  
 complement(4322..4350)  
 /note="factor Sigma70; predicted +1 start at 2642935"  
 complement(4413..6725)  
 /gene="pbpc"  
 /gene="pbpc"

promoter  
 gene  
 CDS

```

/function="putative enzyme; Not classified"
/notes="f770: This 770 aa ORF is 31 pct identical (43 gaps)
to 480 residues of an approx..720 aa protein PPPF_BACSU
SW: P38050"
/transl_table=1
/codon_start=1
/product="putative peptidoglycan enzyme"
/protein_id="AAC75572.1"
/db_xref="GI:1788867"
/translation="MPRLTRKGCWTLTAAPFLFLAAMGADKMLPLHEVNAVY
VVAODGTPLMRFADGDIWRYPTVEDSPRYLALINEEDRWKHPGVPSVARA
AMODLTSGRTVIGSGSTLMQVARDLPHKTFGKIRQLMVALQLEWLSRELITLY
LNRAFGSTLOGIGAASWAYLKSPPANISSEAMLAIVLPAPSLRPDRWPEBABA
RNKYLERAQGVMSRECVKESREPIALAPROMPOLAPLESRLMGSKSDKITTL
DAGLORLEELAAQNMKGRLPPRSILAMIVDHTDMRVGWSVDLDESGHVDVY
NSIRSPSVLKPFIYGLALDDELHPASLLDIDVPRRCQDPRGMPDSGFHGISSEA
LVRSINLPAVONLENGKRFRAKLRNGLPLVINGAPMLSLILGSAKKEDMA
AYTAFARHGKAGKLRLOPDDPLERPLMSGSAWITRIMADEQPLDSALPRVAPL
AMKGTSTGYDANAIGVNAVYVIGIWMRGDIPVVOGFASVPLLNQVNNILLS
RSANLPEDPRNSVTRGVICMPGGSLPEGDGNCRRRLATILDDGSAVPLLEOEG
INGTIRPITLIDNGKRVADCPQARQEMINWPLLEPMLPASERAVRLPASTSCP
PYGHDAQPLDLCYRGDAIIRKLPDGAETLPLDSSGAGERNWFLNGEPLTERGN
VTLRLTDKGDITQLEVMVDYQGIATYKFMQ"
complement(6726..11687)
/gene="b2520"
/complement(6726..11687)
/gene="b2520"
/function="orf; Unknown"
/notes="f1653; f1653"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC75573.1"
/db_xref="GI:1788868"
/translation="MKLRVACMIMLAGCDNNDNAPTAVKNDPSEVTKASSEN
ASSAKLSVPERKILAQOSRGKYLTLIDISEVOLDQAATILVLFSTPLDDQDFSAVH

```

Query Match 100.0%; Score 1119; DB 1; Length 13176;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-270;  
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	Accession	Score	Length	Start	End	Gap
QY	1	1119	13176	1	13176	0
DB	1204	1119	13176	1	13176	0
QY	61	1119	13176	1	13176	0
DB	1144	1119	13176	1	13176	0
QY	121	1119	13176	1	13176	0
DB	1084	1119	13176	1	13176	0
QY	181	1119	13176	1	13176	0
DB	1024	1119	13176	1	13176	0
QY	241	1119	13176	1	13176	0
DB	964	1119	13176	1	13176	0
QY	301	1119	13176	1	13176	0
DB	904	1119	13176	1	13176	0
QY	361	1119	13176	1	13176	0
DB	844	1119	13176	1	13176	0
QY	421	1119	13176	1	13176	0
DB	784	1119	13176	1	13176	0

DB	Accession	Score	Length	Start	End	Gap
QY	481	1119	13176	1	13176	0
DB	724	1119	13176	1	13176	0
QY	541	1119	13176	1	13176	0
DB	664	1119	13176	1	13176	0
QY	601	1119	13176	1	13176	0
DB	604	1119	13176	1	13176	0
QY	661	1119	13176	1	13176	0
DB	544	1119	13176	1	13176	0
QY	721	1119	13176	1	13176	0
DB	484	1119	13176	1	13176	0
QY	781	1119	13176	1	13176	0
DB	424	1119	13176	1	13176	0
QY	841	1119	13176	1	13176	0
DB	364	1119	13176	1	13176	0
QY	901	1119	13176	1	13176	0
DB	304	1119	13176	1	13176	0
QY	961	1119	13176	1	13176	0
DB	244	1119	13176	1	13176	0
QY	1021	1119	13176	1	13176	0
DB	184	1119	13176	1	13176	0
QY	1081	1119	13176	1	13176	0
DB	124	1119	13176	1	13176	0

RESULT 7  
 AY033515 1119 bp DNA linear BCT 26-DEC-2001  
 LOCUS  
 DEFINITION  
 Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 synthase (lspc) gene, complete cds.  
 ACCESSION  
 AY033515  
 VERSION  
 AY033515.1 GI:17978527  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.

REFERENCE  
 1 (bases 1 to 1119)  
 Hecht,S., Eisenreich,W., Adam,P., Amslinger,S., Kis,K., Bacher,A.,  
 Arigoni,D. and Rohdich,F.  
 Studies on the nonmetabolite pathway to terpenes: The role of the  
 Gcpe (lspc) protein  
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)  
 JOURNAL  
 PUBLISHED  
 11752431  
 2 (bases 1 to 1119)  
 Rohdich,F., Hecht,S., Adam,P., Bacher,A. and Eisenreich,W.  
 Direct Submission  
 Submitted (27-Apr-2001) Institut fuer Organische Chemie und  
 Biochemie, Lehrstuhl III, Technische Universitaet Muenchen,  
 Lichtenberg Strasse 4, Garching D-85747, Germany  
 JOURNAL  
 TITLES  
 FEATUERS  
 1. 1119  
 /organism="Escherichia coli"

```

/db_xref="taxon:562"
1..1119
/gene="ispC"
1..1119
/gene="ispC"
/function="converts 2C-methyl-D-erythritol
2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl
4-diphosphate, last but one step of the
mevalonate-independent pathway for isoprenoid
biosynthesis"
/note="ispC"
/codon_start=1
/product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
synthase"
/protein_id="AAK53460.1"
/db_xref="GI:17978528"
/translation="MHNQAPLQRRKSTRIVGNPLIGDGAPIAVOSMTNTRTDVEAT
VNOIKALERYGADIVRSVPTMDAAEFKLIKQVNPPLVADHFDRIALKAETGV
DCLRINFGNGNERIRMAVDCAADKNIPRIQVNASLERDQEKYGEPTQALLES
AMRHVDHEDLNEDQKVSASDVLAHSVRIIAKQIDQPLHGLTEAGARGAV
KSAIGLEELSEGIDPLRVSLADPVEIKVGDILKSLRISRGINFACPTCSHQ
EFVDIGTVNLEORLEDIITPMDSIIGCVNNGEALVSTLGVGTGKKSGILEYEDGV
RKRLDNDMDIDLEARIKAKASOLDEARRIDVQOVER"
BASE COUNT      272 a      280 c      315 g      252 t
ORIGIN

```

Query Match 97.7% Score 1093.4; DB 1; Length 1119;  
 Best Local Similarity 98.6%; Pred. No. 2,7e-264;  
 Matches 1103; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 ATGCATACACGAGGCGCAATTCAGTAAATCAACACCTATTACTGTGGGAATGTG 60
Db 1 ATGCATACACGAGGCGCAATTCAGTAAATCAACACCTATTACTGTGGGAATGTG 60
QY 61 CCGATTGGCGATGTGCTGGCATGCGGTACAGTCATGACCAATAGCGGTACAGAC 120
Db 61 CCGATTGGCGATGTGCTGGCATGCGGTACAGTCATGACCAATAGCGGTACAGAC 120
QY 121 GTGCAAGCAAGCGTCATGAAATCAAGCGGTGGAAGCGGTGGCGGTATATGTCGCT 180
Db 121 GTGCAAGCAAGCGTCATGAAATCAAGCGGTGGAAGCGGTGGCGGTATATGTCGCT 180
QY 181 GTATCCGTACGAGATGAGCGGCGGCAAGCGGTCAAACTCAACAGCAGCATTAAC 240
Db 181 GTATCCGTACGAGATGAGCGGCGGCAAGCGGTCAAACTCAACAGCAGCATTAAC 240
QY 241 GTGCCGCTGTGCTGACATGCACTTCGATTCGATTCGATTCGATTCGATTCGATTC 300
Db 241 GTGCCGCTGTGCTGACATGCACTTCGATTCGATTCGATTCGATTCGATTCGATTC 300
QY 301 GGGCGTGAATGTGCGTAAACCCCGGAATATGCGTAAATGAAGCGGTATTCGATG 360
Db 301 GGGCGTGAATGTGCGTAAACCCCGGAATATGCGTAAATGAAGCGGTATTCGATG 360
QY 361 GTGGTGAATGTGCGGCGGCAAAATTCGATTCGATTCGATTCGATTCGATTCGATTC 420
Db 361 GTGGTGAATGTGCGGCGGCAAAATTCGATTCGATTCGATTCGATTCGATTCGATTC 420
QY 421 CTGCAAAAACATCTGCAAGAAAGTATGCGAAGCGGCGGAGCGGTTCGATTCAT 480
Db 421 CTGCAAAAACATCTGCAAGAAAGTATGCGAAGCGGCGGAGCGGTTCGATTCAT 480
QY 481 GCGATGCGCATGTTGATGATCTGCGATCGCTTAACCTGATAGTTCAAGTACACG 540
Db 481 GCGATGCGCATGTTGATGATCTGCGATCGCTTAACCTGATAGTTCAAGTACACG 540
QY 541 AAAGCGTGAAGCTCTTCTGCGCTGTTAGTCTTATGTTGCGCAAAACGATTCAT 600
Db 541 AAAGCGTGAAGCTCTTCTGCGCTGTTAGTCTTATGTTGCGCAAAACGATTCAT 600
QY 601 CAGCGCTGATCTGGGATGACCGAAGCGGCTGTGCGGCGGCGGCGGCACTMAATCC 660
Db 601 CAGCGCTGATCTGGGATGACCGAAGCGGCTGTGCGGCGGCGGCGGCACTMAATCC 660

```

```

Db 601 CAGCGCTGATCTGGGATGACCGAAGCGGCTGTGCGGCGGCGGCGGCACTMAATCC 660
QY 661 GCGATGTTAGGCTGCTGCTGCTGCAAGCGATCGGCGGCGGCGGCTGCTGCTG 720
Db 661 GCGATGTTAGGCTGCTGCTGCTGCAAGCGATCGGCGGCGGCGGCTGCTGCTG 720
QY 721 GCGGCGGATCGGCTGCAAGAGATCAAGTCCGTTTCGATTTATTTGAATCGCTGATC 780
Db 721 GCGGCGGATCGGCTGCAAGAGATCAAGTCCGTTTCGATTTATTTGAATCGCTGATC 780
QY 781 CGTTGCGAGGATCACTTATTCGCTGCCCGGACCTGTTGCGGTGCGAATTTGATGTT 840
Db 781 CGTTGCGAGGATCACTTATTCGCTGCCCGGACCTGTTGCGGTGCGAATTTGATGTT 840
QY 841 ATCGGTACGTTTACGCGCTGAGCAAGCGCTGAGAGATATCATCTCCGATGAGCTT 900
Db 841 ATCGGTACGTTTACGCGCTGAGCAAGCGCTGAGAGATATCATCTCCGATGAGCTT 900
QY 901 TCGATTATCGGCTCGGTGATGATGCGCCAGGTGAGCGGTGTTTCTACACTCGGCTC 960
Db 901 TCGATTATCGGCTCGGTGATGATGCGCCAGGTGAGCGGTGTTTCTACACTCGGCTC 960
QY 961 ACCGGCGGCAACAGAAAGAGCGGCTCTATGAGATGCGGTGCGCAAAAGCGTGTGAC 1020
Db 961 ACCGGCGGCAACAGAAAGAGCGGCTCTATGAGATGCGGTGCGCAAAAGCGTGTGAC 1020
QY 1021 AACACGATATGATGACACAGCTGAGCAAGCGATTCGTCGAAAGCGCATGAGCTGAG 1080
Db 1021 AACACGATATGATGACACAGCTGAGCAAGCGATTCGTCGAAAGCGCATGAGCTGAG 1080
QY 1081 GAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119
Db 1081 GAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119

```

RESULT 8  
 AE005481/c  
 LOCUS  
 DEFINITION  
 Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 100 of 290.  
 ACCESSION  
 AE005481 AE005174  
 VERSION  
 AE005481.1 GI:12516908  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli O157:H7 EDL933.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
 REFERENCE  
 1 (bases 1 to 11521)  
 Perna N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
 Genome sequence of enterohemorrhagic Escherichia coli O157:H7  
 Nature 409 (6819), 529-533 (2001)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 21074935  
 11206551  
 REFERENCE  
 2 (bases 1 to 11521)  
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
 Direct Submission  
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 1..11521  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"

gene /db\_xref="taxon:155864"  
/note="enterohemorrhagic"  
complement(86..1204)  
/gene="gcpE"  
/note="23778"  
complement(86..1204)  
/gene="gcpE"  
/function="orf: unknown function"  
/note="Residues 1 to 372 of 372 are 100.00 pct identical  
to residues 1 to 372 of 372 from Escherichia coli K-12  
strain MG1655: B2515"  
/codon\_start=1  
/transl\_table=1  
/product="orf, hypothetical protein"  
/protein\_id="AAG57625.1"  
/db\_xref="GI:12516909"  
/translation="MHNOAPQRRKSTRIYGVNPIGDGAPIAVOSMTPTTVEAT  
VNQIKALERVGADIVRSVPTMDAENAFKLIKQOVNPLVADITHEDRIALKAIEGY  
DCLRNPNQNGENERIRNVDCARDKNPIRIGVNASLEKDLQEKTEGPEPQALLES  
AMRHVDHLRLNPDQFVSKASDVFLAVESYRLAKQIDPLHIGITEAGARGAV  
KSAIGLILLEGIGDTLRVSLADPVEIKVGFILKSLRISRGINFACPSRC  
EDVIGTVNALEORLEDIITPMQVSIIGCVVNGPEALVSLGVTGCKKSGIVEDGY  
RKDLNDNDMIDQLEARIKAKASQLEARRIDVOQVER"  
complement(1231..2244)  
/gene="yfgA"  
/note="23779"  
complement(1231..2244)  
/gene="yfgA"  
/function="putative membrane: Not classified"  
/note="Residues 1 to 337 of 337 are 99.10 pct identical to  
residues 1 to 337 of 337 from Escherichia coli K-12 strain  
MG1655: B2516"  
/codon\_start=1  
/transl\_table=1  
/product="putative membrane protein"  
/protein\_id="AAG57626.1"  
/db\_xref="GI:12516910"  
/translation="MTEATHDQNEALTTGARLNARNEQLGSOQAVARLCLKYSTV  
RDIEDKADADLASFELRGYSYARLVHPEEQLPELEKQAPLRKAKVAPMSQESL  
GSRKKRQGMTEFTMLVLFVYIGLSGAMMDDHKAQOEITTMADSSXELSSQ  
GQVPLNTXTTDDRTSTTPASVDTTATNTQTPVTPARAVPQQAIVSPSANY  
DTATPATPATTPDGAAPLPDQAGVTPPADPVALVMTADLWELVTATGKLF  
SGMQRKDNLM/LGQAPYKLRIGAPAAVQIYQCKPVDLSRFIRTNQVARTLNAEOS  
PAQ"  
complement(2529..3683)  
/gene="yfgB"  
/note="23780"  
complement(2529..3683)  
/gene="yfgB"  
/function="orf: unknown function"  
/note="Residues 1 to 384 of 384 are 99.47 pct identical to  
residues 1 to 384 of 384 from Escherichia coli K-12 strain  
MG1655: B2517"  
/codon\_start=1  
/transl\_table=1  
/product="orf, hypothetical protein"  
/protein\_id="AAG57627.1"  
/db\_xref="GI:12516911"  
/translation="MSEOLVTPENTTKDKINILDLNRQOMREFPKDLGKPPRADO  
VAKMYTHCDPDEMDINKYLRGLKVAEIRAEVEEQRSDGTTKMAIVAGDO  
REYVYIPEDDRATLCVSSQVCALECFCSAQOGEFNRNLRVSEIIGQVMAAIVG  
LAKYVGRPITNVVMGEPFLNNVNPAMEIMLDFGGLSKRRYTLSTSGVPA  
DDKTDMDIDVALAISLHAPNDEIRDEIPIKXXKXETFLAVRYLRKSNANOGRT  
IEYVWLHDNGDTEHAHOLAELKDTPECKINLIPNPAPAPYGRSSNSRIDRESKVL  
MSYGTITVIRKTRGDDIDAACQLAGVDYIDRKRLRRKMOGEAIDIKAV"  
complement(3833..4264)  
/gene="ndk"  
/note="23781"  
complement(3833..4264)  
/gene="ndk"  
/function="enzyme: Nucleotide biosynthesis: Purine  
ribonucleotide biosynthesis"  
/note="Residues 1 to 143 of 143 are 100.00 pct identical

misc\_feature  
gene  
CDS  
to residues 1 to 143 of 143 from Escherichia coli K-12  
strain MG1655: B2518"  
/codon\_start=1  
/transl\_table=1  
/product="nucleoside diphosphate kinase"  
/protein\_id="AAG57628.1"  
/db\_xref="GI:12516912"  
/translation="MATERFESIIRKPAVAKNYIGNIFARPEAGFIVGTKMLHTV  
EOARGFAEDHGKPEFDGLVEFMTSPGVIVSLGEMAVORHDLGATNPANLACT  
LRADYADSLTEGHSDSVESAREIATFFGEGEVCPRTR"  
4397..9255  
/note="O-island #106: Region of the EDL933 chromosome not  
homologous to E. coli K-12 MG1655"  
complement(4405..4983)  
/gene="23782"  
complement(4405..4983)  
/gene="23782"  
/function="orf: unknown function"  
/note="No significant matches"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAG57629.1"  
/db\_xref="GI:12516913"  
/translation="MLMHRKGTITAIADLSQAKAMERALLINECLROYCEGMSVY  
SPHSEVAVSRRLPHVQRADVMTQIQPGRLRLHLPQISDRPMDPOKCOLCA  
CMRACEDQVFSLNGHQLINDALCNGCNAVCFOAMTVEIITLPAKIVLHANRK  
VCKTCOKSFTEFOQNAONCLYCCRHRGRMRP"  
complement(5259..6080)  
/gene="23783"  
complement(5259..6080)  
/gene="23783"  
/function="putative enzyme: Not classified"  
/note="Residues 2 to 272 of 273 are 25.26 pct identical to  
residues 6 to 276 of 286 from GenPept 118 :  
g1150021261gb1AADC7317.11AET35170.8 (AF135170) dimethyl  
sulfoxide reductase subunit C [Yersinia pestis]"  
/codon\_start=1  
/transl\_table=1  
/product="putative dimethyl sulfoxide reductase subunit C"  
/protein\_id="AAG57630.1"  
/db\_xref="GI:12516914"  
/translation="MHLEPLAFTLLDASVGMTLAFELRHILITSSAPVVRPL  
LLAPLAGVLLISTLHGYPLNMAHLSHSSSLWSEITFGALYLAJLSTLLVA  
LRKGMOLLMLAAVGVIVDECMAGQVYIHSITTWQYNTLIMFLGVGLGSALVA  
VFRISGILPQIDALRNGCVLIALVLRLLVLPWIMDITLANMAQATLPHAPLML  
GOLPILTLISNGISVIGMFPFVAGCKKNIPALFSGVMLVGSVEMLRVFFSIG"  
complement(6073..6510)  
/gene="23784"  
complement(6073..6510)  
/gene="23784"  
/function="putative enzyme: Not classified"  
/note="Residues 4 to 134 of 145 are 61.83 pct identical to  
residues 57 to 186 of 205 from GenPept 118 :  
g117871221gb1AAC73981.11(AE000191) anaerobic dimethyl  
sulfoxide reductase subunit B [Escherichia coli]"  
/codon\_start=1  
/transl\_table=1  
/product="putative dimethyl sulfoxide reductase subunit B"  
/protein\_id="AAG57631.1"  
/db\_xref="GI:12516915"  
/translation="MSHNVFAYTTLSSVCNHCADPCTKNCPTMMHKKRPGGIVAVNT  
DKVCGCGASDVCGPQAMQMEQGMKSCDFCIDLQAKGQAPCVAVCPGALIFGP  
IDLAREKYGVDVSGGLPDSSTITQPNLVIKPHQGAKEASHHA"  
complement(6699..9080)  
/gene="23785"  
complement(6699..9080)  
/gene="23785"  
/function="putative enzyme: Not classified"  
/note="Residues 25 to 792 of 793 are 43.59 pct identical  
to residues 3 to 784 of 785 from GenPept 118 :  
g117871221gb1AAC73980.11(AE000191) anaerobic dimethyl  
sulfoxide reductase subunit A [Escherichia coli]"  
/codon\_start=1





TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS
Shinagawa, H.				
Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12		21156231	5 (bases 1 to 296827)	Omishiri, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
Direct Submission				
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kengen-info.osaka-u.ac.jp, <a href="http://www.gen-info.osaka-u.ac.jp/">http://www.gen-info.osaka-u.ac.jp/</a> , Fax:81-6-6879-2047)				
genome project:				
Location/Qualifiers				
1. .296827				
/organism="Escherichia coli O157:H7"				
/strain="O157:H7"				
/sub_strain="RIMD 0509952"				
/db_xref="taxon:83334"				
414. .1484				
/gene="ECS3123"				
414. .1484				
/gene="ECS3123"				
/note="unknown"				
/codon_start=1				
/evidence=not_experimental				
/transl_table=11				
/product="hypothetical protein"				
/protein_id="BAB36546.1"				
/db_xref="GI:13362593"				
/translation="MPMTSHLSRCNLTQNNKLTQPTVEIKESQATKLEFFCDEFLKI SOSYRGSTSDSIANLAEFTKASDAISIKQKIDVHSYGEVPIRHFVEHYITKN ERFADVEYTLTPTFRGNENNGEILSVEKEGISTORTNLMEYEAIRNI INEHYIKENI HMYSDPEKYEIKSINSDDKPYKTSKESIKSOLYSDORNIOCYTLRKNTWYKKA SSEERIDFKFNSKSEIKNTYKQIYQAKRLNOMKYIKLSGELIDITDKRLAPLINDS FSVYRDFEAFSKQKERNLFTPDNSKFVPKREKGLMVHNSGLVITGKIQEGVOTAF TOEDYKILIRINIFTEPKLE"				
complement(1708. .2784)				
/gene="ECS3124"				
complement(1708. .2784)				
/gene="ECS3124"				
/note="similar to GLPQ_ECOLI g11788572 percent identity 99 in 358 aa (Conserved in E.coli K-12)"				
/codon_start=1				
/evidence=not_experimental				
/transl_table=11				
/product="glycerophosphodiester phosphodiesterase"				
/protein_id="BAB36547.1"				
/db_xref="GI:13362594"				
/translation="MKTLKLNLSAIMKSTIWMGSSAMAADSENEKIVIAHRGASCLYP EHTLPKAMAYAGADYLEDLVMTKDHVLVLDHVLDRVTDAADRFPPRARDDGKYR EYAIPTLDEIKSLKTEGEGDIENCKKQVYGRPMGSDRFVHTFEEELFVQGLKRIH STGKELGIYPERIKAPMHFHOGSKIDIAKLTLEVKKYGTDKDDVYLQCFADALKRHS KNELEPKKMDLNLVOLIAAYTDNNEGTOOKOPDGSNVNNTMDMRKPGAMOVAYADG IGPBTHLIEETSGPGRNKLITGMVQDAQNKLVVPTVBSDKLPETTYDYNQLYDML YNKAVGVNLTDFDPKAVPLNKE"				
complement(2789. .4147)				
/gene="ECS3125"				
complement(2789. .4147)				
/gene="ECS3125"				
/note="similar to GLPT_ECOLI g11788573 percent identity 99 in 452 aa (Conserved in E.coli K-12)"				
/codon_start=1				
/evidence=not_experimental				
/transl_table=11				
/product="sn-glycerol-3-phosphate permease"				
/protein_id="BAB36548.1"				
/db_xref="GI:13362595"				
/translation="MLSTFKPAPHKARLPAAEIDPTYRLRLMOFLGIFPGYAAVLYLV RNFDAAPYLVEGFSRSGDLGFAFGSILVYGGSTFMGVSQDSNRKRLPPLGLIIR AAAYVILFMGFVMAWSSILAVMEVLLFSLGSGVGMGKWPCCGTWTHMWSQKRRGIVSIV				

gene  
4420. 6048  
/gene="ECS3126"  
CDS  
4420. 6048  
/gene="ECS3126"  
/note="similar to GLP\_A\_ECOLI g11788574, percent identity  
100 in 542 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase  
large subunit"  
/protein\_id="BAB36549.1"  
/db\_xref="GI:13362596"  
/transl\_table=1  
/note="similar to KRT20SQSDV111GGATAGIARDICALRGLVILVERHDIAT  
GATSNHDLHSGARVAVTDAESARECSISEMOLKIRARHCEVPEINLFITLPEDDLS  
FOATIRACEEGEISAEALIDPOQARIIEPPANALIGAVKPDGVDFPLTANMLL  
AKEHVALELTHEVGLIREGATGVGVNRHNLNGETOLAHAPVVVNAAGITMGHIAE  
YADLIRMPAPKAGSLIMDRHNQHVIMRCRPSADILVPGDITSLIGTSLRDVY  
ELIDNRKVAEEDVILIRECEKLAIPMAKTRILRAISGRPLVASDPPSGNNVSGVIT  
LIDHAEKRGDGEFITTTGKGLMTYRLHMEWATDAVCRKLGTRPCTTADLALPSGE  
AEVTRLRVYSIPAPLRGSAVYRNHGRTPAMTESEGLHSLVCECAVTAEGVQAVAEV  
LAVNSLLDLRRTRPVGKGTCCGELCAACAAGILLOFNPNTTSQMSIEOLSTVLNMRKQ  
VPIAMGDLRRESEFTRMVWGICGLKEKQDAL"  
6038. 7297  
/gene="ECS3127"  
/gene="ECS3127"  
/note="similar to GLP\_B\_ECOLI g11788575, percent identity  
98 in 419 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase  
membrane anchor subunit"  
/protein\_id="BAB36550.1"  
/db\_xref="GI:13362597"  
/transl\_table=1  
/note="MKPFIIMGGGLAGLCLGLOLQKHGLRCAIVTRGOSALHFFSSG  
LIDLSLPGCAVTDIHSGLESLRQOAPRHPYSLPGQFVLDLACOALNLESQAQ  
GCSVELPHORITPTICSTWLSPEPVWPLPARKICVGSGLMDPOAHLAAASL  
LIDLAETLAELELPELDVLRNNATEPRAVNIRKFLDNEENRPLLDALIPVANTCEM  
LMPACGLADDDKLRLWLNELRCLPCSLMLPELTPPSVGLNQLQDROFYRQCGVAMP  
DEVKATVCNGVYNEIWRNHNADILPRPEPVLASGSFSGGLVAERNGIRRPILGLL  
VLOTATRGWYKGDFAPOPMQOFGVTTDQTLRPSAGQGTENLFTGISVLGEPDPI  
OGSGGVCVASLHAAQIQARAGGQ"  
7294. 8484  
/gene="ECS3128"  
/gene="ECS3128"  
/note="similar to GLP\_C\_ECOLI g11788576, percent identity  
100 in 396 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase  
K-small subunit"  
/protein\_id="BAB36551.1"  
/db\_xref="GI:13362598"  
/transl\_table=1  
/note="NMDTSPENCITKCTVCTACPYSVRVNPGYGPKQAGDPDEERLRL  
DGLVDEALATKCYNCRCEVCAQSDVKTIDTIOBARAKYDTTPRSLSRNVLSHTDLM  
SVSTPPAPLVNTRTSIKLPYROLIDALAKLIDHRRITLPKPYFCGFRMYRSVAQAOQY  
DOVAFHGCGVANNHPQLGKDLIKVLANAGTQQLSKRCCGVPLIANGFTDKARK  
ALTINVSIRAPVANGKIPVIAITSSTFALRDEYPEVLNVDNKGLDLELTLRWLM  
KLDEGKTLPELPLKLVVYVHTPCHEMKGWTLVTELLRNINIGLELTVLSDCCGTA  
TVFKKKEENPTSOAIGAPLEFRQIEESGMDLVVDCTCKWQIEMSTSCSEHPITLL  
GALA"  
8603. 8677  
/gene="ECS3129"  
gene  
CDS





|||||  
Db 421 CTGAAAAAGATCTCAGAAAAAGATGGCAAGCCGACGCGCTTGCTGGAATCT 480  
Oy 481 GCCATGCGTCATGTGATCATCTCATGCGCCGAACTTCGATCATTTAAAGTACGGTG 540  
Db 481 GCCATGCGTCATGTGATCATCTCATGCGCCGAACTTCGATCATTTAAAGTACGGTG 540  
Oy 541 AAAGGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 541 AAAGGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Oy 601 CAGCGCTGATCTGAGGATCACCAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
Db 601 CAGCGCTGATCTGAGGATCACCAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
Oy 661 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Db 661 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Oy 721 GCGGCGCATCCGCTGAGAGATCAAGAGTGGTTTGAATTTGAATTCGCTGCTATC 780  
Db 721 GCGGCGCATCCGCTGAGAGATCAAGAGTGGTTTGAATTTGAATTCGCTGCTATC 780  
Oy 781 CATTGCGAGGATCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Db 781 CATTGCGAGGATCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Oy 841 ATCGGTACGTTAAAGCGCTGAGACAGCGCTGGAAGATATCATCTCCGATGAGCTT 900  
Db 841 ATCGGTACGTTAAAGCGCTGAGACAGCGCTGGAAGATATCATCTCCGATGAGCTT 900  
Oy 901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Db 901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Oy 961 ACCGCGGCGACACAGAAAAGCGGCTCTATGAAAGATGCGCTGCGCAAGA 1010  
Db 961 ACCGCGGCGACACAGAAAAGCGGCTCTATGAAAGATGCGCTGCGCAAGA 1010

RESULT 11  
AR208328 1010 bp DNA linear PAT 20-JUN-2002  
LOCUS AR208328 Sequence 8 from patent US 6383745.  
DEFINITION AR208328  
ACCESSION AR208328  
VERSION AR208328.1 GI:21509454  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1010)  
AUTHORS Rather,P.N.  
TITLE Methods of screening for anti-microbial utilizing aarc and  
JOURNAL compositions thereof  
PATENT: US 6383745-A 8 07 -MAY- 2002;  
FEATURES  
Location/Qualifiers  
source 1..1010  
BASE COUNT 238 a 253 c 287 g 232 t  
ORIGIN

Query Match 90.3%; Score 1010; DB 6; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 2.6e-243;  
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ATGCATTAACAGGCTCCATTCATGATGATAAAATCAACAGCTATTACTGTGGGAATG 60  
Db 1 ATGCATTAACAGGCTCCATTCATGATGATAAAATCAACAGCTATTACTGTGGGAATG 60  
Oy 61 CCGATTGGGAGATGCTGCCATCGCGCTACAGTCATACCAATACGCGTACGACAGAC 120  
Db 61 CCGATTGGGAGATGCTGCCATCGCGCTACAGTCATACCAATACGCGTACGACAGAC 120

Oy 121 GTGAAAGCAAGGTCATCAATCAATCAAGGCGCTGGAACCGCTTGCGCTGATATGCTCGT 180  
Db 121 GTGAAAGCAAGGTCATCAATCAATCAAGGCGCTGGAACCGCTTGCGCTGATATGCTCGT 180  
Oy 181 GTATCCGTACCGAGATGAGCGCGGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAC 240  
Db 181 GTATCCGTACCGAGATGAGCGCGGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAC 240  
Oy 241 GTGCGCGTGGTGGGATCATCATCTTATGATGATGCTGCTGCTGCTGCTGCTGCTG 300  
Db 241 GTGCGCGTGGTGGGATCATCATCTTATGATGATGCTGCTGCTGCTGCTGCTGCTG 300  
Oy 301 GCGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
Db 301 GCGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
Oy 361 GTGTTGACTGTGCGCGGATTAACATTCGATTCGCTATTTGGCGTTAAAGCGGATG 420  
Db 361 GTGTTGACTGTGCGCGGATTAACATTCGATTCGCTATTTGGCGTTAAAGCGGATG 420  
Oy 421 CTGAAAAAGATCTGCAAGAAAGATGAGCGAAGCGCGGACGCGCTGCTGCTGATCT 480  
Db 421 CTGAAAAAGATCTGCAAGAAAGATGAGCGAAGCGCGGACGCGCTGCTGCTGATCT 480  
Oy 481 GCCATGCGTCATGTGATCATCTCATGATCGCTGAACTTCGATCATGCTCAAGCTG 540  
Db 481 GCCATGCGTCATGTGATCATCTCATGATCGCTGAACTTCGATCATGCTCAAGCTG 540  
Oy 541 AAAGCGTCTGACGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 541 AAAGCGTCTGACGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Oy 601 CAGCGCTGATCTGAGGATCACCAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
Db 601 CAGCGCTGATCTGAGGATCACCAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
Oy 661 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Db 661 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Oy 721 GCGGCGCATCCGCTGAGAGATCAAGAGTGGTTTGAATTTGAATTCGCTGCTATC 780  
Db 721 GCGGCGCATCCGCTGAGAGATCAAGAGTGGTTTGAATTTGAATTCGCTGCTATC 780  
Oy 781 CATTGCGAGGATCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Db 781 CATTGCGAGGATCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Oy 841 ATCGGTACGTTAAAGCGCTGAGACAGCGCTGGAAGATATCATCTCCGATGAGCTT 900  
Db 841 ATCGGTACGTTAAAGCGCTGAGACAGCGCTGGAAGATATCATCTCCGATGAGCTT 900  
Oy 901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Db 901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Oy 961 ACCGCGGCGACACAGAAAAGCGGCTCTATGAAAGATGCGCTGCGCAAGA 1010  
Db 961 ACCGCGGCGACACAGAAAAGCGGCTCTATGAAAGATGCGCTGCGCAAGA 1010

RESULT 12  
AE008814/c 23647 bp DNA linear BCT 31-JUL-2002  
LOCUS AE008814 Salmomella typhimurium LT2, section 118 of 220 of the complete  
DEFINITION genome.  
ACCESSION AE008814 AE006468  
VERSION AE008814.1 GI:16421058  
KEYWORDS  
SOURCE Salmomella typhimurium LT2.  
ORGANISM Salmomella typhimurium LT2.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Salmomella.

REFERENCE 1 (bases 1 to 23647)  
AUTHORS McCelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,  
Lattelle,P., Courgeon,L., Portolick,S., Ali,J., Dante,M., Du,F.,  
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,  
Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Wilson,L., Miller,W.,  
Stonking,T., Nham,M., Waterston,R. and Wilson,R.K.  
TITLE Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
LT2  
JOURNAL Nature 413 (6858) 852-856 (2001)  
MEDLINE 21534948  
PUBMED 11677609  
REFERENCE 2 (bases 1 to 23647)  
AUTHORS The Salmonella typhimurium Genome Sequencing Project.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of  
Genetics, Washington University School of Medicine, 4444 Forest  
Park Boulevard, St. Louis, MO 63108, USA  
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated  
computer analysis using similarity information and the programs:  
GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and  
GeneMark: <http://opal.biology.gatech.edu/GeneMark/>  
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto  
Encyclopedia of Genes and Genomes: <http://www.genome.ad.jp/kegg/>,  
and Pedro Romero and Peter Karp at EcoCyc:  
<http://ecocyc.pangeasystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites  
were kindly provided by Heladia Salgado, Julio Collado-Vides and  
Reguondb:  
[http://kinloh.cifn.unam.mx:8850/db/reguondb\\_intro.frameset](http://kinloh.cifn.unam.mx:8850/db/reguondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate  
chemistries or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one m3 subclone.

## FEATURES

## source

location/qualifiers

1..23647

/organism="Salmonella typhimurium LT2"

/strain="9827; SGSC 1412; ATCC 700720"

/db\_xref="taxon:700720"

/db\_xref="taxon:99287"

/note="LT2"

complement(110..5711)

/gene="rda"

/note="STM2515"

complement(110..5707)

/gene="rda"

/note="rda (g14583529)"

/codon\_start=1

/transl\_table=11

/product="putative outer membrane protein"

/protein\_id="AAL21409.1"

/db\_xref="GI:16421059"

/translation="MDNRQNVKVVHFLITLIMPAVSTAPAAALLKGTGNOELNSVT  
GAVNGAPLADGAILPLVYOGSTLIDSKTHDEFSAMPDFDSADTSTMRVNSTDT  
EDLSDPZLIMENKOPPAKGLVWADATPTPLSPQVNLITCAQWLAROLVAM  
AGVEDTVEALMLFTRGVNPAIPLSPKVALNKRKATDPTVSDHDAFEA  
SKVGESELTISTKACDTPGNAPEVIRKDAENRQGVNNAVPGVDELTET  
OTREGVANGNATVYVTKGEGKPLVSSVNEPALTAETAVIFETITSPSDK  
ASMGHMTESATATNGITVETTRPLAASGADKSVDTMETVALTFWISADNHC  
ILPDAEOLVOMHEHSTLETWGWPAAGAEVWSKNOISGHAHVHNSASVRAA  
DSDTLVSQVKAQPAHPQITLSPGPKAVGESITLTVVDRDQKPLPYRME  
LEIDAKNKRGHODAMNLRTVDSSEKASPEHITGVTVNGAHLTLKHNKSLG  
VETPIRIYEDDEGNELSFVITVYSPVDGANMGMHGVVADAGNLTKRLPLA  
VASHKGGSEENEMATENSVASATACGAGVGDOSLAHLVSEHAGGOMESHG  
WPTEDYFIADSDASGVAVNLNGDSGCTPTPTLTSANEMAVAVLDVYNDPVI  
KNAADKVGQVAKNHSRNLNGLMAGNTDFTTANGRKRDGLTGFDTSGEMOF  
DDVGYAGGVHGTIDANGATITLTKRGVGLTLPNIAVDSLITPVSXKFTV

## CDS

## gene

## CDS

ATSPDPAAMGMHMDITVGDMPFERPLAGEVNSNPLRTODESNETVARVHSDAV  
GNDGAGCANRLPRIDOLEALYNANSGK IHS ICGMPYLANWSTYOSATWKLIA  
LNGSFRFANSNPIYVASCANSPVASTITIEVNSQWYDGGDVAVVKGKGTETOL  
KTVKQVSNPIPEAPFVLRGDIYRGRKRTAOUAGDDLOGLVTVVLDGESLATTT  
KMGSGDTGDTGRILISVTRPDHGTATNALTYENAAVASIDTLETVYVDSGAVAR  
RMGNHAPSLTADGAVOYRPLLYAEIASSDQENETAEVHFGVASECANAR  
CAAGYPAVALDLSYKSPSRTINFAQGMPIYYSVSGNSSTFSSGAKLDEYVAD  
LADGSRSENSATSTAMOYICATPPLPOTATLTSPOAMDALIOVAKNESITPL  
LITTPDAMGPVPAATFESIKRAGKARNNDPNTATSSIPVETVLTSPDSKA  
SVITGATGAGDTALTLAEFGIGLKNOLTANLNDPTTSSIPVETVLTSPDSKA  
NMYGHPETFTASNGAEFRKPLVAGESSSEAHVDYFETENWIMVNSFNTGVS  
YMOAMAIIDETALYNHDPGSKATVDIGLVGKRWAGDSILKOSTLYMOYKDKTKGN  
YSSENDGNYIOLCLTTSRSLINIALSDANADKSAVAKGETIIPMTVYVNDAG  
OPQAGVAVLTRYAYSRCVADKOYIEPVGITGPPVPTSPANMMLAPVAPGTAFA  
NNONGTSKWSGCTGDDKRLPTLTDKSLGKTSTALANOPEASVADLFTVOT  
SPDTPIASTWGHMPDVOYVGLRPTLKAEISAPRDTWPNFNEFWGTNTYIOSEH  
VETSLHLCQSDENIASLDLALQSVITGLQWPTTSSMDVYSDGQSNKYCSFNE  
TTGQTTCTREKATTSGLGCRVP"  
complement(5706..5711)  
/gene="rda"

## RBS

## gene

## gene

## CDS

## CDS

## RBS

## RBS

## gene

## gene

## CDS

## CDS

## RBS

## RBS

## gene

## gene

## RBS

## RBS

## gene

## gene

## CDS

## CDS

ATSPDPAAMGMHMDITVGDMPFERPLAGEVNSNPLRTODESNETVARVHSDAV  
GNDGAGCANRLPRIDOLEALYNANSGK IHS ICGMPYLANWSTYOSATWKLIA  
LNGSFRFANSNPIYVASCANSPVASTITIEVNSQWYDGGDVAVVKGKGTETOL  
KTVKQVSNPIPEAPFVLRGDIYRGRKRTAOUAGDDLOGLVTVVLDGESLATTT  
KMGSGDTGDTGRILISVTRPDHGTATNALTYENAAVASIDTLETVYVDSGAVAR  
RMGNHAPSLTADGAVOYRPLLYAEIASSDQENETAEVHFGVASECANAR  
CAAGYPAVALDLSYKSPSRTINFAQGMPIYYSVSGNSSTFSSGAKLDEYVAD  
LADGSRSENSATSTAMOYICATPPLPOTATLTSPOAMDALIOVAKNESITPL  
LITTPDAMGPVPAATFESIKRAGKARNNDPNTATSSIPVETVLTSPDSKA  
SVITGATGAGDTALTLAEFGIGLKNOLTANLNDPTTSSIPVETVLTSPDSKA  
NMYGHPETFTASNGAEFRKPLVAGESSSEAHVDYFETENWIMVNSFNTGVS  
YMOAMAIIDETALYNHDPGSKATVDIGLVGKRWAGDSILKOSTLYMOYKDKTKGN  
YSSENDGNYIOLCLTTSRSLINIALSDANADKSAVAKGETIIPMTVYVNDAG  
OPQAGVAVLTRYAYSRCVADKOYIEPVGITGPPVPTSPANMMLAPVAPGTAFA  
NNONGTSKWSGCTGDDKRLPTLTDKSLGKTSTALANOPEASVADLFTVOT  
SPDTPIASTWGHMPDVOYVGLRPTLKAEISAPRDTWPNFNEFWGTNTYIOSEH  
VETSLHLCQSDENIASLDLALQSVITGLQWPTTSSMDVYSDGQSNKYCSFNE  
TTGQTTCTREKATTSGLGCRVP"  
complement(5706..5711)  
/gene="rda"

/note="putative RBS for rda; Reguondb: STMSI1H002622"

complement(5827..6793)

/gene="sin1"

/note="STM2516"

complement(5827..6786)

/gene="sin1"

/note="sin1 (g14583530)"

/codon\_start=1

/transl\_table=11

/product="putative outer membrane protein"

/protein\_id="AAL21410.1"

/db\_xref="GI:16421060"

/translation="MQATVRRILRTFVATLAVVAGCAAPAVANAGNSGMOIVSE  
TGTIOCTVMTTAAKATADTKDKHVTYIDGDKRTYEGSKQFHVGRKQVYVNAIG  
DTEGDLDTNITATKATVNSFSQDQSGPKDUGDSYETIOAADRTIGTKITPT  
TGGDPAVATELLKIDLSTDAAGSDDEIPEBQVVDENVHVVYIHKDSHNLKSGT  
TKTNTTVOVLMSKNGTNGTAGENYVDQDYRWKFEVGTGKINGTGTTGTVNESWN  
DKDIVIPVYNAAKTAFEAGDGVTVSGVGGFISIDYKRR"

complement(6786..6793)

/gene="sin1"

/note="putative RBS for sin1; Reguondb: STMSI1H002623"

complement(6844..9042)

/gene="sinH"

/note="STM2517"

complement(6844..9036)

/gene="sinH"

/note="similar to Escherichia coli intimin and Yersinia  
pestis invasins proteins; (g14583531)"

/codon\_start=1

/transl\_table=11

/product="sinH"

/protein\_id="AAL21411.1"

/db\_xref="GI:16421061"

/translation="WVRIRILVLLVSGVYVAAAGDNGSASANDSHSLPDLASES  
ARKEQENKSKFEKGADYFINSATQGDNLTPALSSQARSYLQNDITSSQSYLE  
GVMSPYKIRTSLSVGEQGLDSSLDIIPYDQOSTLFEQISAKRKEEDTINFG  
LGVRONVGMWLLGNAFYDYDTRGHRRLGIDETAMTVDLKSQVNYHPLSKMAGT  
PDYFERRPARGMWLLGNAFYDYDTRGHRRLGIDETAMTVDLKSQVNYHPLSKMAGT  
LEYTVPLVTVGDTKAKGSDNSVNAVTVYQICTPLAADLPENKLOHSLMGR  
TDFVDRNNITILEYREKDPDVTLMWLKADATNEHECVIEDPEAVGLEKRWYNA  
LINTKHYKIIASWQAKNNAARTLVAVYKAKLITDQNNNSMWLVPANWMDTEBORT  
ALNTKVRVETLDEDEGNKONGSVVEITVOODKIELIYDNLADTRSDSHSASALAD  
GSDGVAMDLITDSDGSDNRNGENLVDDAMPVLYDSNDKVTLAQDPTCTETRCV  
IASRKEAGTVVLTSLTGTPEWRKAKADAGSYVVDYFTGDNLSALNAVYQVAA  
NPNVILGKEDKIPYNNNAVRLTLMEDKKDQFQSGQDTEEBALTYIOMFTQOST  
NGHTGALANTMEDLVLPVTNKEAOKFAANVEDSGVIGIKRVTYSOK"

complement(9037..9042)

/gene="sinH"

/note="putative RBS for sinH; Reguondb: STMSI1H002624"

complement(9382..9615)

/gene="yfgJ"

/note="STM2518"

complement(9382..9603)

OY	361	GTGGTTACATCTGGCGCGATATAAAACATTCCGATTCGATTAATGGCGTAAACGCCGAGATC	420
Db	15249	GTTGGTAGCTCGCGCGCGAACAANAACATCCCTATTGCTATCGGGCTAAACGCCGAGATCG	15190
OY	421	CTTGAAAAAAGATCTGCAGAAAAGATATGGCGAACCCGACGCCGACGGCTGCTGGAACT	480
Db	15189	CTTGAAAAAAGATCTGCAGAAAAGATACGGTATAGCCGACGCCGACGGCTGCTGGAACTCA	15130
OY	481	GCCATGCGCATGTTTGATCATCTCGATCGCCTGAACCTTCGATTCAGTTCAAGTTCAGCGCTG	540
Db	15129	GCCATGCGCCCTATGTCGATCATCTCGATCGCTGCAACTTGTATCAAGTTCAAGTTCAGCGCTG	15070
OY	541	AAAGCGTCTGACGCTTCCTCGCTGTTGAAGTCTTATCGTTTGGTCGAAAAACAGATCGAT	600
Db	15069	AAAGCGTCTGATGATATTCCTGCGCTTGAAGTCTTATCGTTTGGTCGAAAAACAGATCGAT	15010
OY	601	CAGCGCTGCATCTGGGAGATCACCGAAGCCGGTGTGCGCGACGCGGGCGAGTAAATTC	660
Db	15009	CAGCGCTGCATCTGGGAGATCACCGAAGCCGGGCGCGCTAGCGGGGCGGTCAAGTTCG	14950
OY	661	GCCATGCTTAAAGTCTGCTGCTGTCTGAAGAAGCATGGGCGACGCTCGGGTATCGGTG	720
Db	14949	GCGATCGGTTTAAAGTCTGCTGCTGTCTGAAGAAGTATGGCGACGCTCGGGTATCGGTG	14890
OY	721	GCGCGCGATCCGGTTCGAGAAGATCAAAGTCGGTTTCGATTTTGGAAATTCGTCGGCTATC	780
Db	14889	GCGCGCGATCCGGTTCGAGAAGATCAAAGTTCGGTTTCGATTTTCGAAAGTTCGTCGGCAT	14830
OY	781	CGTTCCGAGAGGATCACTTCATCGCTGCGCCCGACCTGTTGCGCGTCAGAAATTTATGTT	840
Db	14839	CGCGCGCGCGGATCAATTTATCCGCTGCGCGACCTGTTTCGCGCAGACATTCACAGCT	14770
OY	841	ATCGGTACGGTTTAAAGCGCGCTGGAGCAAGCGCTCGAAGAAATATCATCTCCGATGGACGTT	900
Db	14769	ATCGGACAGGGTGAAGCGCGTGGAGCAGCTCTGGAAATATCATCAACCCTATGGACGTC	14710
OY	901	TGCAATTATCGGCTGCGTGGTGAATGAGCCACAGGTGAAGCGCTGCTTCTTAACACTCGCGCTC	960
Db	14709	TGCAATCATCGTGTGGCTATGTAAGCGGGCGGGCGAAGCGCTGATTCAGACGCTGGGCGTG	14650
OY	961	ACCGCGCGACACAGAAAAGCGCCTCTATATGAATATGGCGTCCGCAAAAGCCGTCGAGAC	1020
Db	14649	ACGGCGCGCAATTAAGAAAAGCGCCTGTATGAAGACGGCGTCCGTATAGACAGACTCGAT	14590
OY	1021	AACACGATATGATGACACAGCTGGAAGCAGCATTCGTGGCAAAAGCCGTCAGCGAGAC	1080
Db	14589	AACGACGATATGATGCGCAGCTGTGAATCCGTTATTCGCGCAAAAGCAAGTCAACTTGAT	14530
OY	1081	GAAGCGCGTCGAATTGACGCTTCAGCAGAGTTGAAAAATAA	1119
Db	14529	GAAGCGCGTCGAGATTGACGCTGCTGCAGGTGAAAAATAA	14491

RESULT 13  
AL627275/c  
LOCUS  
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,  
complete chromosome, segment 11/20.  
ACCESSION AL627275 AL513382  
VERSION  
KEYWORDS  
SOURCE Salmomella enterica subsp. enterica serovar Typhi.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Salmomella.

REFERENCE  
AUTHORS 1 (bases 1 to 145050)  
Parikh,I.D., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,  
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,  
Sebahia,M., Baker,S., Basham,D., Davies,R.M., Dowd,L., White,N.,  
Comerton,P., Cronin,A., Davis,P., Davies,K., Chillingworth,T.,  
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holtzoy,S.,  
Jagals,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Garra,P.,

TITLE Stevens, K., Whitehead, S. and Barrell, B.G.  
Complete genome sequence of a multiple drug resistant Salmonella  
JOURNAL enterica serovar Typhi Ctr18  
MEDLINE Nature 413 (6858), 848-852 (2001)  
PUBMED 21534947  
REFERENCE 11677608  
AUTHORS 2 (bases 1 to 145050)  
TITLE Parkhill, J.  
JOURNAL Direct Submission  
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella  
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK  
E-mail: parkhill@sanger.ac.uk  
Notes:  
Details of S. typhi sequencing at the Sanger Centre are available  
on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).  
FEATURES  
source  
1. 145050  
/organism="Salmonella enterica subsp. enterica serovar  
Typhi"  
/strain="Ctr18"  
/db\_xref="taxon:90370"  
636..640  
/note="possible RBS"  
647..1003  
/gene="STY2720"  
/note="yifB"  
647..1003  
/gene="STY2720"  
/note="Orthologue of E. coli yifB (YFEB\_ECOLI); Pasta hit  
to YFEB\_ECOLI (118 aa), 81% identity in 118 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAD02682.1"  
/db\_xref="GI:16503699"  
/db\_xref="SPTRMBL:0824S3"  
/translation="MTLVGIRKNDITIKARRMLEHGIDYRHDYRVGDIDPLPLNT  
FLAEIWMQPLNRTGRTWRKLDGHRGSLTDADSAALMEQPAIIKRPLICAPGRFM  
LIGFESRYQOFDEV"  
995..1000  
/gene="STY2720"  
/note="possible RBS"  
1007..2134  
/gene="STY2721"  
/note="dape"  
1007..2134  
/gene="STY2721"  
/note="Orthologue of E. coli dape (DAPE\_ECOLI); Pasta hit  
to DAPE\_ECOLI (375 aa), 94% identity in 375 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="succinyl-diaminopimelate desuccinylase"  
/protein\_id="CAD02683.1"  
/db\_xref="GI:16503700"  
/db\_xref="SPTRMBL:0824S2"  
/translation="MSCPYELTQQLIRPSPSPDDVGCALMIERLRIKIGTIEHND  
EGDTYNVAMARRGEGTFLAAGHTDVVPAGVDWMINPPEPTIRDMGLFGRAADKKG  
SLAMVAAERFAQHPHRRGLAFLITSDDEASAKNGTVKVEALMARERLDYCLV  
GEPSTIEIVGVKNGRSGSLTNCNLTIHGOGVAVPHLADNPHRAPLINLVIE  
MDRGNDPEPATSMOINAGTGSNNVIGELFOPNFNPSTETDPMIKERHALLE  
KHQRLTYDMMWLSGCPFLTARGLVAVVNAIENYNEIKPOLITGTGSDGRITAKMG  
AOYVELGPVNAITHKINECVNADQLLARMTORIMEQLVA"  
1031..1978  
/gene="STY2721"  
/note="Pfam match to entry PF01546 Peptidase M20.  
Peptidase family M20/M25/M40, score 315.30, E-value  
7.4e-91"  
1187..1216  
/gene="STY2721"  
/note="PS00758 Arge / dape / ACY1 / CPG2 / yscs family  
signature 1"

misc-feature  
1295..1414  
/gene="STY2721"  
/note="PS00759 Arge / dape / ACY1 / CPG2 / yscs family  
signature 2"  
2150..2153  
/note="possible RBS"  
2162..2362  
/gene="STY2722"  
2162..2362  
/gene="STY2722"  
/note="Similar to Escherichia coli similar to TR\_P77259  
(EMBL:D90875) (66 aa) fasta scores: E(): 1e-27, 95.5% id  
in 66 aa. Contains hydrophobic, possible membrane-spanning  
region"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAD02684.1"  
/db\_xref="GI:16503701"  
/db\_xref="SPTRMBL:0824S2"  
/translation="MDLAKYMWLVLVFLVGLVNLVYIKLRIDHKKFLANKPELP  
HRDENRKMDEDDMPKKDOPK"  
complement(2389..4404)  
/gene="STY2723"  
complement(2389..4404)  
/gene="STY2723"  
/note="Orthologue of E. coli YPFI\_ECOLI; Pasta hit to  
YPFI\_ECOLI (671 aa), 72% identity in 662 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAD02685.1"  
/db\_xref="GI:16503702"  
/db\_xref="SPTRMBL:0824S1"  
/translation="MSDIDALRALTSOMTQEGIRLLVTSQAWCERARAVLVP  
GMMVAPDAPROPCRTPOALQTLIGREPHNATFDANQGDAAFAALSTLQASWLL  
LMPETYESRPTDLSLRNSDCAQPIPTPQKQHLKRLSPQTLNRQRPFCWP  
STPSRWRPATEGPQEQAILSLRLREMPGVAVYIARGRKSAIAQFISOMGT  
AIVTAPAKATADILIAFAGEFCEFPDALLASGARWLVVDEAAIAPILQIYS  
RPFRIILTTTVOGTEGRTGFLKFCARFPQHLRFTLRQVMAPECPLENTYSELI  
FDEEFAQAPDHGIIASAFYQOAMGKTPALPRAVYQLLSGANYRTSPDLRRMDAG  
OHLQATANNRVAALMLVEEGLSAELSAVWCGPRRGNVQSTIAHGSPLAA  
TIVGRVSTIATVHARQREGISQQLIACCEQAQCDYISVFCTTPELRRWRQEG  
VIVRWGSHREASGCTAMALLPLSDAGRLAQOERRRLRQADILITWNGEATPLAA  
LREQALNDEDWRELWGFARHRLTSLCLRLLOYSAIPLPLAGLEKASDAEL  
CARLIRISGRKALLALRAQAAQALIALDAGRTRQLRDVMPGCGDHAG"  
complement(2878..3207)  
/gene="STY2723"  
/note="Pfam match to entry PF00583 Acetyltransf,  
Acetyltransferase (GNAT) family, score 32.50, E-value  
9.7e-06"  
complement(3790..3813)  
/gene="STY2723"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement(4420..5283)  
/gene="STY2724"  
complement(4420..5283)  
/gene="STY2724"  
/note="Orthologue of E. coli P76563; Pasta hit to P76563  
(287 aa), 92% identity in 286 aa overlap. Contains  
hydrophobic, possible membrane-spanning region"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAD02686.1"  
/db\_xref="GI:16503703"  
/db\_xref="SPTRMBL:0824S0"  
/translation="NRWGRRESNNVEDRRNRPGBSLGPGFRLPRGKGITILVVV  
VVAAYGVDLTLTGQVPSQOSTRSISPNDDEAKFTSVLTATEDTWGLFKMG  
RGYQOPKLVYWRGMRRTGCGAGQSVWGPYCADQTVYIDLSFYVMNKGADGDF  
OGVYIAHEVGHVOKLIGTEPKVROLQOASQTEVNRISVREMLADCEPAGVWGHSMO  
OGGVLEAGDLEELNAAQAIGDRLIQGQGGQGVYVDSFTHGHSBQRIYSFKRKGPSGD  
PACNTFGKNP"

```

RBS      complement(5287..5292)
          /note="possible RBS"
          /complement(5415..6128)
          /gene="STY2725"
          /note="STY2725"
          /note="purc"
          /complement(5415..6128)
          /gene="STY2725"
          /note="Orthologue of E. coli purc (PUR7_ECOLI); fasta hit
          to PUR7_ECOLI (237 aa), 95% identity in 237 aa overlap"
          /codon_start=1
          /transl_table=11
          /product="phosphoribosylaminoimidazole-succinocarboxamide
          synthase"
          /protein_id="CAD02687.1"
          /db_xref="GI:16503704"
          /db_xref="SPTREMBL:O8XF69"
          /translation="MOKAELKRGKAKVYSTENDLLVLEPRNDSAGDARIEOPD
          RGVNANKFNHIMTKLAAGIPLOMERLSDTELVKLEKVPYECVRRNRAQSLV
          KRLVEEGMELNPLIEDLFKNDALHDMVNSVCEYTGWVQENLARKELTYKAND
          VLKLFDDAGLIDFKLEFGLYKGEVVLGDEFSRSLMDKETLMDKDRFROSL
          GGLIYAVAVAHRLGCKLD"
          /complement(5421..6122)
          /gene="STY2725"
          /note="Pfam match to entry PF01259 SAICAR synt, SAICAR
          synthetase, score 414.20, E-value 1.2e-120"
          /complement(5586..5612)
          /gene="STY2725"
          /note="PS01058 SAICAR synthetase signature 2"
          /complement(5829..5873)
          /gene="STY2725"
          /note="PS01057 SAICAR synthetase signature 1"
          /complement(6137..6141)
          /note="possible RBS"
          /complement(6303..7337)
          /gene="STY2726"
          /note="nlpB"

Query Match      79.1%; Score 885.4; DB 1; Length 145050;
Best local Similarity 87.0%; Pred. No. 7.8e-212;
Matches 973; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY      1 ATGCATACCAGGCTCCATTCATCAAGTAGAATAATCAACAGCTATTACGTGGGAATG 60
        |||||||
DB 64624 ATGCATACCAGGCTCCATTCATCAAGTAGAATAATCAACAGCTATTACGTGGGAATG 64565

QY      121 GTGGAAGCAACGGTCAATCAATCAAGGCGCTGGAAGCGCGTGGCGCTGATATCGTC 180
        |||||||
DB 64504 GTGGAAGCAACGGTCAATCAATCAAGGCGCTGGAAGCGCGTGGCGCTGATATCGTC 64445

QY      181 GTATCCGTCACCGAGATGACGCGCGCAAGAGCGTTCAAACTCATCAAAACGCAAGTTAAC 240
        |||||||
DB 64444 GTTTCGTACCGACATGACGCGCGCAAGAGCGTTCAAACTCATCAAAACGCAAGTTAAC 64385

QY      241 GTGCGCGTGGTGGTGACATCCACTTCGATTCGATTCGCTGCGTGAAGTAGCGAATAC 300
        |||||||
DB 64384 GTGCGCGTGGTGGTGACATCCACTTCGATTCGATTCGCTGCGTGAAGTAGCGAATAC 64325

QY      301 GGGGTGCTATTGTCGCTATTAAACCTGGAATATCGGTAAAGAGCGTATTCCGATG 360
        |||||||
DB 64324 GGGGTGCTATTGTCGCTATTAAACCTGGAATATCGGTAAAGAGCGTATTCCGATG 64265

QY      361 GTGCTTACCTGTGCGCGCGCAATAACATTCGATTCGCTGCTTAAAGCGCGGATTCG 420
        |||||||
DB 64264 GTGCTTACCTGTGCGCGCGCAATAACATTCGATTCGCTGCTTAAAGCGCGGATTCG 64205

QY      421 CTGGAATAAAGATCTGCAAGAAAGTATGAGAACCGACGCGGAGCGGTGCTGGAATCT 480
        |||||||
DB 64204 CTGGAATAAAGATCTGCAAGAAAGTATGAGAACCGACGCGGAGCGGTGCTGGAATCT 64145

```

```

QY      481 GCCATGCGTCATGTTGATCATCTGATCGCGCTGAACCTTCGATCACTTCAAGTCAGCGTG 540
        |||||||
DB 64144 GCCATGCGCGCGATGTCATCATCTGATCGCGCTGAACCTTCGATCACTTCAAGTCAGCGTG 64085

QY      541 AAAGCGTCTGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
        |||||||
DB 64084 AAAGCGTCTGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 64025

QY      601 CAGCGCTTCATCTGCGGAGATCACCGAAGCGCGTGGCGCGCGAGCGGGGCGAGTAATAATCC 660
        |||||||
DB 64024 CAGCGCTTCATCTGCGGAGATCACCGAAGCGCGTGGCGCGCGAGCGGGGCGAGTAATCC 63965

QY      661 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
        |||||||
DB 63964 GCCATTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63905

QY      721 GGGGCGGACCGCGTGAAGAGATCAAGAGCGGCTTCGATTTTGAATTCGCTGCGCTATC 780
        |||||||
DB 63904 GGGGCGGACCGCGTGAAGAGATCAAGAGCGGCTTCGATTTTGAATTCGCTGCGCTATC 63845

QY      781 CGTTGCGAGGAGATCACTTCATTCGCTGCGCGGACCTGCTTCGAGGAATTTGATGTT 840
        |||||||
DB 63844 CGGCGCGCGCGGATCAATTTATTCGCTGCGCGGACCTGCTTCGAGGAATTTGATGTT 63785

QY      841 ATGCGTAGCTTTAAGCGCGCTGCGAGCAAGCGCTGGAAGATATCATCTCCATGAGACGTT 900
        |||||||
DB 63784 ATGCGTAGCTTTAAGCGCGCTGCGAGCAAGCGCTGGAAGATATCATCTCCATGAGACGTT 63725

QY      901 TCGATTATCGGCTGGTGTGTAATGCGCGAGCGCGCTGCTTCATCTGCGCGCTC 960
        |||||||
DB 63724 TCGATTATCGGCTGGTGTGTAATGCGCGAGCGCGCTGCTTCATCTGCGCGCTC 63665

QY      961 ACCGCGGCAACAGAAAGCGCGCTTATGAAGATGCGCGCAAGACCGCTGCGAC 1020
        |||||||
DB 63664 ACCGCGGCAACAGAAAGCGCGCTTATGAAGATGCGCGCAAGACCGCTGCGAC 63605

QY      1021 AACACGATATGATGACGACGCTGGAAGCAAGCGCTTTCGCAAGACCGCTGAGTGC 1080
        |||||||
DB 63604 AACACGATATGATGACGACGCTGGAAGCAAGCGCTTTCGCAAGACCGCTGAGTGC 63545

QY      1081 GAAGCGCGTGAATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1140
        |||||||
DB 63544 GAAGCGCGTGAATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 63506

RESULT 14
D90881/c
LOCUS
DEFINITION
E.coli genomic DNA, kohara clone #428(56.8-57.0 mln.).
ACCESSION
D90881.1 GI:1799913
VERSION
Complete and shotgun sequencing; PBPE; POM; gcpe; ndk; yfgA; yfgB.
KEYWORDS
Escherichia coli (strain:K12) DNA, clone_11b:kohara lambda miniset
SOURCE
library clone:kohara clone #428.
ORGANISM
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Yamamoto,Y., Alba,H., Baba,T., Hayashi,K., Inada,T., Isono,K.,
Itoh,T., Kimura,S., Kitagawa,M., Makino,K., Miki,T., Mitsuhashi,N.,
Mizobuchi,K., Mori,H., Nakade,S., Nakamura,Y., Nishimoto,H.,
Oshima,T., Oyama,S., Saito,N., Sempel,G., Setoh,Y., Takemoto,K.,
Sivasubraman,S., Tagami,H., Takahashi,H., Takeda,Y., Takemoto,K.,
Uehara,K., Wada,C., Yamagata,S., and Horiiuchi,T.
Construction of a contiguous 874-kb sequence of the Escherichia
coli -K12 genome corresponding to 50.0-68.8 min on the linkage map
and analysis of its sequence features
DNA Res. 4 (2), 91-113 (1997)
JOURNAL
MEDLINE
97349980
REFERENCE
2 (sites)
Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiiuchi,T.,
Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,

```





QY 181 GTATCGGTACCGACGATGACCGGACGAGAGGTTCAACATCATCAACAGCAGGTTAC 240  
 |||||||  
 DB 530 GTATCGGTACCGACGATGACCGGACGAGAGGTTCAACATCATCAACAGCAGGTTAC 471  
 QY 241 GTGGCGCTGGTGGTACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 300  
 |||||||  
 DB 470 GTGGCGCTGGTGGTACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 411  
 QY 301 GGCGTCGATTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 360  
 |||||||  
 DB 410 GGCGTCGATTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 351  
 QY 361 GTGGTCGATTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 420  
 |||||||  
 DB 350 GTGGTCGATTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 291  
 QY 421 GTGGAAGAAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 480  
 |||||||  
 DB 290 GTGGAAGAAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 231  
 QY 481 GCCATCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 540  
 |||||||  
 DB 230 GCCATCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 171  
 QY 541 AAGCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 600  
 |||||||  
 DB 170 AAGCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 111  
 QY 601 CACCGCTGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 660  
 |||||||  
 DB 110 CACCGCTGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 51  
 QY 661 GCCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 710  
 |||||||  
 DB 50 GCCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1  
 RESULT 15  
 D90880/c 16214 bp DNA linear BCT 29-MAY-1997  
 LOCUS D90880  
 DEFINITION E.coli genomic DNA, Kohara clone #427(56.5-56.9 min.).  
 ACCESSION D90880 AB001340  
 VERSION D90880.1 GI:1805564  
 KEYWORDS Complete and shotgun sequencing; AFSK; gcpe; guab; huss;  
 SOURCE xseq  
 Escherichia coli (strain:K12) DNA, clone\_lib:Kohara lambda miniset  
 library clone:Kohara clone #427.  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 REFERENCE 1 (sites)  
 AUTHORS Yamamoto, Y., Baba, T., Fujita, K., Hayashi, K., Inada, T., Isono, K.,  
 Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsuhashi, N.,  
 Mizobuchi, K., Mori, H., Nakada, S., Nakamura, Y., Nishimoto, H.,  
 Oshima, T., Oyama, S., Saito, N., Sempel, G., Satoh, Y.,  
 Sivasubramanian, S., Tagami, H., Takahashi, H., Takeda, J., Takemoto, K.,  
 Uehara, K., Wada, C., Yamagata, S. and Horiuchi, T.  
 Construction of a contiguous 874-kb sequence of the Escherichia  
 coli -K12 genome corresponding to 50.0-68.8 min on the linkage map  
 and analysis of its sequence features.  
 JOURNAL DNA Res. 4 (2), 91-113 (1997)  
 MEDLINE 97349980  
 REFERENCE 2 (sites)  
 AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T.,  
 Ikenoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,  
 Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,  
 Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,  
 Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,  
 Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,  
 Yamamoto, Y. and Yano, M.  
 TITLE The systematic sequencing of the Escherichia coli genome in Japan  
 JOURNAL Unpublished

REFERENCE 3 (bases 1 to 16214)  
 AUTHORS Mori, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUN-1997) Hirotsada Mori, NARA Institute of Science  
 and Technology, Res. & Edu. Center for Genetic Info.: 8916-5  
 Takayama, Ikoma, Nara 630-01, Japan  
 (E-mail:hmori@glc.ais-t-nara.ac.jp, Tel:81-7437-2-5660,  
 Fax:81-7437-2-5669)  
 COMMENT Collaboration Information:  
 Project:  
 The Japan E. coli genome DNA sequencing project  
 Group:  
 The Japan E. coli genome DNA sequencing group  
 Members: (1995.4 - 1996.3)  
 Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,  
 Ikenoto, K., Inada, T., Isono, K., Isono, S.,  
 Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,  
 Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,  
 Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,  
 Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,  
 Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,  
 Yamamoto, Y. and Yano, M.  
 Headed by:  
 Name: Takashi Horiuchi  
 Address: National Institute of Basic Biology, Okazaki, 444, Japan  
 E-mail: kishori@nib.ac.jp  
 Information operator:  
 Name: Hirotsada Mori  
 Address: NARA Institute of Science and Technology,  
 Ikoma, 630-01, Japan  
 E-mail: hmori@glc.ais-t-nara.ac.jp  
 URL:  
 The Japan E. coli genome database  
 http://sw3.ais-t-nara.ac.jp.  
 Location/Qualifiers  
 1. 16214  
 /organism="Escherichia coli"  
 /strain="K12"  
 /db\_xref="taxon:562"  
 /map="56.5 min"  
 /clone="Kohara clone #427"  
 /note="Nucleotide position 2619941-2636154 from the  
 initiation site of TNA (0 min.). This clone is from  
 Kohara lambda miniset library"  
 <1..1494  
 /note="similar to [PIR Accession Number A45333]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="exopolyphosphatase ppx"  
 /protein\_id="BA16392.1"  
 /db\_xref="GI:1805565"  
 /translation="DLGNSFHMVAVRVGAMQITGRKORVHLADGIDPDMNSE  
 AMTRGKNCISLFAERLOGFSPASVCIIGVHTLRALNATIDFLKRAEVIPIPIEITIG  
 NEARFLGKGVGHTDPEKRLVIDIGSTGLVGENEPELVSRKGVSPFOLY  
 PEGGVINKENPRAMAAOKETLTMOPRLOGMVAVGSGTIKAHEVLMEMEKD  
 GITTEPRGLKVEKVEPRHQRQVRSFASLAAQYHDSQARVLDITTKQTEQREO  
 OKRLAIPQLEALIRMAAMKEVGLNINSHGLRHSAVILQNSDLGPFNOQOLMAVTL  
 VYHRAIKLIDLPRTLEKFKQPLDILRLGYVLNNORQATTPPLTLITDOSH  
 WTLRPHDWSNALVLDLEKEQEWEGVAGMRILKEESTPEIAA"  
 complement(1533..3776)  
 /note="similar to [SwissProt Accession Number Q04855]"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BA16393.1"  
 /db\_xref="GI:1805566"  
 /translation="MKLNATYIKIRDMKMGDLPFLPSLILFPAHINTPEAHISGEVF  
 LTRPLALMISMMPFSNATGICATGIVRKYAEIGFETISLANFIITIIICMG  
 YRVEYTRNNVSHGDRLLSQRIQMOIVPAPLILFLPAPAVGLAKAREVNGMP  
 FNLGILNIOALLVNLIGVPLCTIIRVNLNDFYLRSTYSQIKQOVQDLAKREKAL  
 WLIALGALLILCMPLNEKSTFSTFVTLSDLLPLMAMGRYGRKLISLMAVVMIL  
 SIHSYQNIPIYPIGYTQTLITSSSYLVSFIVNYMAVLATRORAVRRIORLAVVP

gene	complement (5796. .7373) /gene="guua" /note="similar to [PIR Accession Number A24640]" /codon_start=1 /transl_table=11 /product="GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2)" /protein_id="BA16394.1" /db_xref="GI:1805567" /translation="MTENIHKRIILIDFGSQYTOLVARKRELGVYCELMANDVTEA QIRNFPSGIIILSGPESTEENSPPAPVFPVGVCGQMDMOMOLGADTEA SNEBEFGYAOVENVDSALVRIEDALTADGKPLDVMHSGDKTAISDEFTVAST ESCFPALMAEERKRFYGVOPHPEVTHROGRMLERVRDITCCCEALMPKATITDVA ARIRBOGDKVTILGSGVDSVTAMLHRAIGKMLTCEVONGLRLNEEYOVDAM FCGHGLNTHVAEDPFLSALGENDPEKRRIIIGRVPEVDEBALLEDVKMLAO GTTFPDVIESAASATKRAHVIKSHHNVGGLPKMKMGVLEPLKEFLKEDVRKIGLELG LPPDVIYRHFPGPLGVRLGEVKEKCYCDLRRADAIIEELIRKADIDVKYQATVY FLPVRSYGVGDGRKYDWMVSLRAVETIDPMTHMHLHYDPLGRSNMILNEVNSIS RYVYDISGRPATIEME"
CDS	complement (7442. .8977) /gene="guab" complement (7442. .8977) /gene="guab" /note="similar to [PIR Accession Number A93555]" /codon_start=1 /transl_table=11 /product="IMP dehydrogenase (EC 1.1.1.205)" /protein_id="BA16395.1" /db_xref="GI:1805568"
gene	transl_id="M05V568" /translation="MOSVTLICIMPROYLLTTEVEILPMILIAKEALTEDVLLVPAHS TVLNRADISTQTLTKTIRLNLPMLSAAMOTVEARLALAOEGGEGFHKHMSIERO AEEVRNADJESQVTVTPDTLPTTLREVKELTENFGFAPVTEENELGCTTIGR DVEVPLDKNRPSVYVTPKRELVYVEGASREYVLAHMEKREYKALVYDDEFLHIG IYKVDOKRANRPNACDKOGRIRVGAAGAGAGNDERVDALVAAGVDLLIDSSHGH SEVYLORIRETRAKTIPDLIIGGNATTAAGALAEAGCSAVKVIIGPISCTTIVT GSPVEPLTVAEAVALLEGSTLIPVLAADGIRPSDGIARKIAGAASAVWGMSLACTEE SPEELIYGGRSKSYFQSMGSLGAMSGSSDRYFQSDNADKLVEGLEGGRAVYKGRLL KEIHHOQMGLRSCMGLTGCCTIDELTRKAEFVRISGAGIQSHVDVLTITESPVNR LGS"
CDS	9070. .10440 /gene="xsea" 9070. .10440 /gene="xsea" /note="similar to [PIR Accession Number A25940]" /codon_start=1 /transl_table=11 /product="exodeoxyribonuclease VII (EC 3.1.11.6) large chain"
gene	protein_id="BA16396.1" /db_xref="GI:1805569" /translation="MLPSOSPAIFTVSRLNOTVRLLLEHMGOWMISGEIENFOPAS GHMYVTLDDTDOVRCAFM RNSNRVTRFPPOGCOVALRANITLVEPGADVOITVEM OPAAGLLOOKKEOLKALIOAGLEFDIOYKPLPSPACHVANTITSGEAAHLHLHLV KRKDSLPVITYIPAVOGDDAPGOIVRIELANORNEDEVILVGRGSLDMLSFND ERYVAITTSRTPVSAVGHETDITAFVADLRAPVPSAAAEVVSRRMOQLLEOVOS TRORLEAMADTYLANRTRFTQIHRLQOQHPOURLAQOQMRLOKRMSEFALENOL KRTGGQOQRLTORLNOQNPORIKHRAQRIQOLEVRLAETLRADLSAREFGNAVTH LEAVSPLESTLARGSVTTATDGNVLRKQYKAGEMLTTRLEDGMISEVKNIQPVK SRKKVH"
CDS	complement (10722. .12233) /note="similar to [SwissProt Accession Number P44536]" /codon_start=1 /transl_table=11

Query Match	61.7%; Score 690; DB 1; Length 16214;
Best Local Similarity	100.0%; Pred. No. 1e-162;
Matches	690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
gene	/protein_id="BA16397.1" /db_xref="GI:1805570" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYKRR LANYFRKSLIDVMSGSP IRIQFKEGENDPNYKRNITLPTPOKRRKRLMKHIKKNK"
CDS	complement (12312. .13490) /gene="AFSK" complement (12312. .13490) /gene="AFSK" /note="similar to [SwissProt Accession Number P54741]" /codon_start=1 /transl_table=11 /product="SERINE/THREONINE PROTEIN KINASE AFSK (EC 2.7.1.1)"
gene	/protein_id="BA16398.1" /db_xref="GI:1805571" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYKRR LANYFRKSLIDVMSGSP IRIQFKEGENDPNYKRNITLPTPOKRRKRLMKHIKKNK"
CDS	complement (12312. .13490) /gene="AFSK" complement (12312. .13490) /gene="AFSK" /note="similar to [SwissProt Accession Number P54741]" /codon_start=1 /transl_table=11 /product="SERINE/THREONINE PROTEIN KINASE AFSK (EC 2.7.1.1)"
gene	/protein_id="BA16399.1" /db_xref="GI:1805572" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYKRR LANYFRKSLIDVMSGSP IRIQFKEGENDPNYKRNITLPTPOKRRKRLMKHIKKNK"
CDS	complement (12312. .13490) /gene="AFSK" complement (12312. .13490) /gene="AFSK" /note="similar to [SwissProt Accession Number P54741]" /codon_start=1 /transl_table=11 /product="SERINE/THREONINE PROTEIN KINASE AFSK (EC 2.7.1.1)"
gene	/protein_id="BA16400.1" /db_xref="GI:1805573" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYKRR LANYFRKSLIDVMSGSP IRIQFKEGENDPNYKRNITLPTPOKRRKRLMKHIKKNK"
CDS	complement (12312. .13490) /gene="AFSK" complement (12312. .13490) /gene="AFSK" /note="similar to [SwissProt Accession Number P54741]" /codon_start=1 /transl_table=11 /product="SERINE/THREONINE PROTEIN KINASE AFSK (EC 2.7.1.1)"
gene	/protein_id="BA16401.1" /db_xref="GI:1805574" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYKRR LANYFRKSLIDVMSGSP IRIQFKEGENDPNYKRNITLPTPOKRRKRLMKHIKKNK"
CDS	complement (12312. .13490) /gene="AFSK" complement (12312. .13490) /gene="AFSK" /note="similar to [SwissProt Accession Number P54741]" /codon_start=1 /transl_table=11 /product="SERINE/THREONINE PROTEIN KINASE AFSK (EC 2.7.1.1)"
gene	/protein_id="BA16402.1" /db_xref="GI:1805575" /translation="MKCLMITYKNEALNMVYVALGRPNVSKTLFNLTRTPDALVA DPGGLTRDREKYGARLEGEFFICIDTGGIDGTEDEVRMAOSDLEADYVLFWMY DRAGIMPEDEALIAHLSREKEFPTIDVANKTGDDPDPAVDVFLSGLGEIPIAASH GGVSTLEHVLTPMMEADLPOEVEDOEAERVAOPEAESEDEDEDFDQSLPIKAI AYGRNPNRKSITLNRILIGERVYVYDMEPTTRRSITIPMERDEHYLIDTAVGRK KAITDVEKFSVITKQILQIEDANVMVYIDAEKISDLSLGLTILNSGSLIVYVA KWDGSLQEVKEQVKTLDLRLGIDFIAVHFTLSALHGSGVNLFESEVADSSSTRVY GTSMLTRIMTNAVEDHOEPLVLRGRVRKLVYAHAGVNP IYVHNQVKKDLPDSSYK

Mon May 12 08:44:41 2003

us-09-921-992-3.rge

Page 21

Search completed: May 10, 2003, 16:14:16  
Job time : 3624 secs

---



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 14:31:07 ; Search time 320 Seconds

(without alignments)  
7874.957 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataaccgctcccat.....ttcagcaggttgaaataaa 1119

Scoring table: IDENTITY-NUC

Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq\_101002:\*

1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	21	AAA95478
2	1119	100.0	1119	21	AAAB8705
3	1119	100.0	1119	22	AA82653
4	1119	100.0	1119	24	AA031203
5	622	55.6	1830121	17	AAT42063
6	411.8	36.8	5484	23	AA573006
7	403.6	36.1	640681	24	ABA92787
8	382.6	34.2	33140	22	AAR28536
9	356	31.8	1857	23	AA590066

10	293.4	26.2	1083	24	ABK74401
11	254.6	22.8	6157	24	ABO70939
12	251.4	22.5	1107	24	ABO67966
13	251.4	22.5	1107	24	ABO69993
14	244	21.8	35829	23	AA559573
15	243.8	21.8	1164	22	AAH52056
16	243.6	21.8	994	24	ABO68324
17	239.4	21.4	1134	22	AAH67170
18	239.4	21.4	349980	22	AAH68530
19	239.4	21.4	349980	22	AAH68531
20	211.8	18.9	1180	19	AAK14017
21	136.4	12.2	507	21	AACT5580
22	136.4	12.2	507	24	ABN76496
23	88	7.9	1152	23	AA588480
24	87.2	7.8	2535	20	AAK20674
25	72.4	6.5	4467	10	AAK92428
26	70.6	6.3	584	24	AAK31220
27	68.4	6.1	596	24	AAK31222
28	68.4	6.1	670	24	AAK31221
29	58.8	5.3	60873	21	AAK81469
30	58.8	5.3	349980	21	AAK21610
31	56.8	5.1	348	24	ABK78811
32	53.4	4.8	1038602	20	AAZ01425
33	52.6	4.7	293	24	AAK31224
34	52.6	4.7	456	24	AAK31225
35	52.6	4.7	504	16	AAQ99779
36	50.8	4.5	1398	20	AAK20609
37	50	4.5	403	24	AAK31223
38	48.8	4.4	564	24	AAK31234
39	48.4	4.3	379	24	AAK31215
40	48.4	4.3	528	24	AAK31214
41	48.4	4.3	2520	24	AAK31201
42	48.2	4.3	353	24	AAK31238
43	48.2	4.3	464	24	AAK31232
44	47.2	4.2	601	24	AAK31210
45	45.4	4.1	613	24	AAK31231

## ALIGNMENTS

RESULT 1	AAA95478	standard; DNA; 1119 BP.
ID	AAA95478	
XX	AAA95478;	
AC	27-FEB-2001 (first entry)	
DT		
XX	E. coli essential gene gcpe.	
DE		
XX	Bacterial growth; inhibitor; ygbB; yfnc; yacB; ychB; yejD; yrfI;	
KW	yggT; yjeE; yjaO; yrcG; yhbC; ygbP; yjeY; gcpe; kdsB; pfs; yaaI;	
KW	b180b; yeaA; yagF; b1983; yldD; yocG; yjbc; antibacterial compound;	
KW	H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;	
KW	M. tuberculosis; antibiotic; ds.	
XX	Escherichia coli.	
OS		
XX	WO200061793-A2.	
PN		
XX	19-OCT-2000.	
PD		
XX	07-APR-2000; 2000MO-EP03135.	
PF		
XX	09-APR-1999; 99EP-0107031.	
PR	04-FEB-2000; 2000EP-0102111.	
XX	(GPCB-) GPC BIOTECH AG.	
PA		
XX	Loferer H, Jacobl A;	
PI		
XX	WPI; 2000-687048/67.	
DR		

Bacillus lichenif  
Listeria monocytog  
Listeria monocytog  
Listeria monocytog  
Propionibacterium  
Mycobacterium tube  
Listeria monocytog  
C glutamicum codin  
C glutamicum codin  
H. pylori GHO 76  
Human ORF1135  
Human ORF143 CDNA  
DNA encoding novel  
Polynucleotide seq  
Sequence encoding  
Zea mays partial g  
Zea mays partial g  
Zea mays partial g  
N. meningitidis pa  
Neisseria meningit  
Bacillus clausii g  
Complete genome se  
Zea mays partial g  
Zea mays partial g  
Cross-reactive al  
Polynucleotide seq  
Zea mays partial g  
Glycine max partia  
Arabidopsis thalia  
Arabidopsis thalia  
Glycine max partia  
Pinus taeda partia  
Zea mays partial g

Query Match	Best Local Similarity	100.0%	Score 1119;	DB 21;	Length 1119;
Matches 1119;	Conservative	0;	Mismatches	0;	Indels
					Gaps
xx	Identifying antibacterial compounds, comprises identifying an				
pt	antagonist or inhibitor of the expression of a gene encoding a				
pt	polypeptide essential for bacterial growth or survival				
xx					
xx	Claim 1; Fig 1; 75pp; English.				
ps					
xx					
cc	The present invention relates to antagonists and inhibitors of 24				
cc	bacterial genes and proteins. The proteins are thought to be essential				
cc	for growth in several species of bacteria (including <i>S. pneumoniae</i> , <i>B.</i>				
cc	<i>burgdorferi</i> , <i>H. influenza</i> and <i>H. pylori</i> ). The proteins and coding				
cc	sequences shown in the specification can be used to identify antagonists				
cc	and inhibitors which can be used in disease treatment and pesticides. In				
cc	particular, they can be used against <i>M. tuberculosis</i> . The present				
cc	sequence is one of the genes of the invention.				
xx					
sq	Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;				
Qy	Query Match	100.0%	Score 1119;	DB 21;	Length 1119;
Db	Best Local Similarity	100.0%	Pred. No. 0;		
Qy	Matches 1119;	Conservative	0;	Mismatches	0;
Db					
Qy	1 ATGCATTAACCCAGGCTCCCAATCAAGTAGAATAATCAACCGTATTACGTTGGGAATGTC				
Db	1 ATGCATTAACCCAGGCTCCCAATCAAGTAGAATAATCAACCGTATTACGTTGGGAATGTC				
Qy	61 CCGATTGGCGATGAGTGTGCTGGCATCCCGGACAGTCCATATACCAATACGGTACGACAGAC				
Db	61 CCGATTGGCGATGAGTGTGCTGGCATCCCGGACAGTCCATATACCAATACGGTACGACAGAC				
Qy	121 GTGCAAGCAACAGCGTCAATGTAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGTCGT				
Db	121 GTGCAAGCAACAGCGTCAATGTAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGTCGT				
Qy	121 GTGCAAGCAACAGCGTCAATGTAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGTCGT				
Db	121 GTGCAAGCAACAGCGTCAATGTAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGTCGT				
Qy	181 GTATCCGTAACGACGATGAGAGCGGCGCAGAAAGCGTTCAAACTCATCAACACAGAGGTTAC				
Db	181 GTATCCGTAACGACGATGAGAGCGGCGCAGAAAGCGTTCAAACTCATCAACACAGAGGTTAC				
Qy	241 GTGCCGCTGGTGGCTGACATATCCACTTTCGACTATTCGCTATTCGCTGAAGTAGCGGAATAC				
Db	241 GTGCCGCTGGTGGCTGACATATCCACTTTCGACTATTCGCTATTCGCTGAAGTAGCGGAATAC				
Qy	301 GGGCGTCAATGTCGTCGTAATTAACCCCTGGCAATATCGGTAATGAAGAGCGTATTTGGCATG				
Db	301 GGGCGTCAATGTCGTCGTAATTAACCCCTGGCAATATCGGTAATGAAGAGCGTATTTGGCATG				
Qy	361 GTGCTTGAAGTGTGCGCGGATAAAAACAATTCGATCCGTATGAGCGTTAACCCCGCATG				
Db	361 GTGCTTGAAGTGTGCGCGGATAAAAACAATTCGATCCGTATGAGCGTTAACCCCGCATG				
Qy	421 CTGGAATAAAGATCTGCAAGAAAGATATGCGCAACCGACGCGGCGTGTGGAATCT				
Db	421 CTGGAATAAAGATCTGCAAGAAAGATATGCGCAACCGACGCGGCGTGTGGAATCT				
Qy	481 GGCATGCGTCAATGTTGATCACTCGATGCGCTGCAACTTCGATCTTCAAAATCGCGCG				
Db	481 GGCATGCGTCAATGTTGATCACTCGATGCGCTGCAACTTCGATCTTCAAAATCGCGCG				
Qy	541 AAAGCGTCTGACGCTTCGCGGCTGTGATGCTTATATGCTTTGCTGGCAAAACAGATCGAT				
Db	541 AAAGCGTCTGACGCTTCGCGGCTGTGATGCTTATATGCTTTGCTGGCAAAACAGATCGAT				
Qy	601 CAGCGCTTGCAATCGGGGATGACGGAACCGGTGGTGGGCGACAGCGGGCAGTAATAATCC				
Db	601 CAGCGCTTGCAATCGGGGATGACGGAACCGGTGGTGGGCGACAGCGGGCAGTAATAATCC				
Qy	661 GCATTTGGTTAGTGTCTGCGCTGCTGTAAGAGCAATCGCGACAGCGTCCGCGTATCGCTG				
Db	661 GCATTTGGTTAGTGTCTGCGCTGCTGTAAGAGCAATCGCGACAGCGTCCGCGTATCGCTG				
Qy	721 GCGGCGCATCCGGTTCGAAGAGATCAAAAGTCGTTTCGATATTTGAAATCGCTGCGTATC				
Db	721 GCGGCGCATCCGGTTCGAAGAGATCAAAAGTCGTTTCGATATTTGAAATCGCTGCGTATC				

QY	781	CGTTCGAGGAGATCACTTCATCGCCTGCGCAGCCGTTCGGTAGGAATTTGATGT	840
Db	781	CGTTCCGGAGGATCACTTCATCGCCTGCGCAGCCGTTCGGTAGGAATTTGATGT	840
QY	841	ATCGGTACGGTTAAACGCGCTGCGAGCAACGCGCTGGAAGATATCATCTCCGATGGACGT	900
Db	841	ATCGGTACGGTTAAACGCGCTGCGAGCAACGCGCTGGAAGATATCATCTCCGATGGACGT	900
QY	901	TGCAATTATCGGCTGCGTGTGTGAATGGGCCAGGTGAGGCGCTGGTTCTTACACTCGAGCGTC	960
Db	901	TGCAATTATCGGCTGCGTGTGTGAATGGGCCAGGTGAGGCGCTGGTTCTTACACTCGAGCGTC	960
QY	961	ACCGGCGGCAACAAGAAAACCGGCTCTATGAAGATGGCGTGGCGAAACCGCTTGGAC	1020
Db	961	ACCGGCGGCAACAAGAAAACCGGCTCTATGAAGATGGCGTGGCGAAACCGCTTGGAC	1020
QY	1021	AACACAGATATGATCGACAGCTGGAAAGCAGCATTCGTGGCGAAAGCCGATGACGCGGAC	1080
Db	1021	AACACAGATATGATCGACAGCTGGAAAGCAGCATTCGTGGCGAAAGCCGATGACGCGGAC	1080
QY	1081	GAAGCGCGTGAATTGACGCTTCACAGACGAGTTGAAAAATAA	1119
Db	1081	GAAGCGCGTGAATTGACGCTTCACAGACGAGTTGAAAAATAA	1119

CC	XX	RESULT 2
CC	XX	ID AAA88705
CC	AC	AAA88705 standard; DNA; 1119 BP.
CC	DT	05-FEB-2001 (first entry)
CC	DE	E. coli FUN essential gene gcpE.
CC	KW	FUN gene; gcpE gene; essential gene; antibacterial; antibiotic;
CC	KM	screening; Infection; therapy; antagonist; surrogate marker; ds.
CC	OS	Escherichia coli.
CC	XX	EP1043403-A1.
CC	PN	11-OCT-2000.
CC	PD	09-APR-1999; 99EP-0107031.
CC	PE	09-APR-1999; 99EP-0107031.
CC	PR	09-APR-1999; 99EP-0107031.
CC	XX	(GPCG-) GPC GENOME PHARM CORP AG.
CC	PA	Not given;
CC	PI	WPI; 2000-640125/62.
CC	DR	
CC	PT	Identifying antagonists of the expression of gene encoding bacterial
CC	PT	growth polypeptide useful for treating bacterial infections or
CC	PT	diseases, by evaluating transcription of the gene in the presence of
CC	PT	test molecule -
CC	XX	
CC	PS	Claim 1; Page 25; 55pp; English.
CC	XX	The present sequence is that of the Escherichia coli FUN gene
CC	XX	gcpE, which encodes a protein that is essential for bacterial
CC	XX	growth or survival. gcpE is 1 of 22 E. coli genes (see
CC	XX	AAA88692-713) identified as being essential (there is no deletion
CC	XX	genotype). These 22 genes fulfill criteria for being attractive
CC	XX	antibacterial targets: hypothetical open reading frames coding for
CC	XX	essential functions (mutation is lethal for growth in rich media);
CC	XX	broad conservation (orthologues are present in a wide range of
CC	XX	bacteria, including Haemophilus influenzae, Streptococcus pneumoniae
CC	XX	Helicobacter pylori and Borrelia burgdorferi); and low toxicity
CC	XX	potential in higher organisms (mostly no orthologues were identified





CC potential antibacterial, antimycotic, antiparasitic or antiviral agents  
CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
CC or herbicidal agents for agriculture.  
XX

SO Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Query Match 100.0%; Score 1119; DB 22; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACACCGGCTCCATTCACGTAAGAAATCAACACGATATTACGTTGGAAATGTG 60  
DB 1 ATGATACACCGGCTCCATTCACGTAAGAAATCAACACGATATTACGTTGGAAATGTG 60  
QY 61 CCGATTGGCGATGGTGTCCCATCGCCGTACAGTCATGACCAATACGCGTACGACAGAC 120  
DB 61 CCGATTGGCGATGGTGTCCCATCGCCGTACAGTCATGACCAATACGCGTACGACAGAC 120  
QY 121 GTGAGACCAACGCTCAATCAATCAAGCGCTGAGACGCGTTGGCGGTGATATGTCGGT 180  
DB 121 GTGAGACCAACGCTCAATCAATCAAGCGCTGAGACGCGTTGGCGGTGATATGTCGGT 180  
QY 181 GTATCCGACCGAGATGAGGCGGCGAGAGCGCTTCAAACTCAACACGACGATTAAC 240  
DB 181 GTATCCGACCGAGATGAGGCGGCGAGAGCGCTTCAAACTCAACACGACGATTAAC 240  
QY 241 GTGCGCGTGTGTGACATCACTTGCATGATGCAATGCGGTGGAAGTACGCGAATAC 300  
DB 241 GTGCGCGTGTGTGACATCACTTGCATGATGCAATGCGGTGGAAGTACGCGAATAC 300  
QY 301 GCGCTCATGTTGCTGCTATTAAACCTTCCGATTCGATGATGAAGACGATATCGCATG 360  
DB 301 GCGCTCATGTTGCTGCTATTAAACCTTCCGATTCGATGATGAAGACGATATCGCATG 360  
QY 361 GTGCTTACTGTGGCGCGATTAACATCCGATTCGATTCGATTCGATTCGATTCGATTCG 420  
DB 361 GTGCTTACTGTGGCGCGATTAACATCCGATTCGATTCGATTCGATTCGATTCGATTCG 420  
QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCAGCGCGCGCTTGTCTGGAATCT 480  
DB 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCAGCGCGCGCTTGTCTGGAATCT 480  
QY 481 GCGATGCGTATGTTGATCATCTGATCGCTGATCACTTGCATCAAGTACGCGTGT 540  
DB 481 GCGATGCGTATGTTGATCATCTGATCGCTGATCACTTGCATCAAGTACGCGTGT 540  
QY 541 AAAGCGTACAGCTCTCCGCTGCTGATGATCTGATGATGATGATGATGATGATGATG 600  
DB 541 AAAGCGTACAGCTCTCCGCTGCTGATGATCTGATGATGATGATGATGATGATGATG 600  
QY 601 CAGCGCTGATCTGGGGATCACCGAAGCGGTGTGCGCGACGCGGCGAGTAAATCC 660  
DB 601 CAGCGCTGATCTGGGGATCACCGAAGCGGTGTGCGCGACGCGGCGAGTAAATCC 660  
QY 661 GCGATTGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
DB 661 GCGATTGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
QY 721 GCGGCGATCCGCTGCAAGAGATCAAAAGTGGTGTGATTTGAAATGGCGGATAC 780  
DB 721 GCGGCGATCCGCTGCAAGAGATCAAAAGTGGTGTGATTTGAAATGGCGGATAC 780  
QY 781 CGTTGCGAGAGATCAACTTCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 840  
DB 781 CGTTGCGAGAGATCAACTTCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 840  
QY 841 ATCGGTAACGCTTAACGCGCTGAGCAACGCGCTGAGATATCACTCCGATGAGCGTT 900  
DB 841 ATCGGTAACGCTTAACGCGCTGAGCAACGCGCTGAGATATCACTCCGATGAGCGTT 900  
QY 901 TCGATTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 901 TCGATTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

QY 961 ACCGCGGCAACAGAAAGCGCTCTATGAGATGCGCGGCAAGACCGCTGAGAC 1020  
DB 961 ACCGCGGCAACAGAAAGCGCTCTATGAGATGCGCGGCAAGACCGCTGAGAC 1020  
QY 1021 AACACGATATGATGACACGACCTGGAAGCAACGATGCTGCGAAGACCGCTGAGAC 1080  
DB 1021 AACACGATATGATGACACGACCTGGAAGCAACGATGCTGCGAAGACCGCTGAGAC 1080  
QY 1081 GAACGCGGTGATGACGCTTACAGCAGTTGAAATTA 1119  
DB 1081 GAACGCGGTGATGACGCTTACAGCAGTTGAAATTA 1119

RESULT 4  
AAD31203  
ID AAD31203 standard; DNA; 1119 BP.  
XX  
AC AAD31203;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Escherichia coli gcpE gene.  
XX  
KM gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;  
KM transgenic plant; isoprenoid compound; tocopherol; isopentenyl diphosphate;  
KM food; feed source; transfection; single nucleotide polymorphism; SNP;  
KM oxidative stress tolerance; UV tolerance; transformation; plant; ds.  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT 1..1119  
FT CDS /tag= a  
FT /product= "Escherichia coli GCPe protein"  
PN WC200212478-A2.  
PD 14-FEB-2002.  
PX  
PF 06-AUG-2001; 2001MC-US24335.  
PR 07-AUG-2000; 2000US-223483P.  
PA (MONS ) MONSANTO TECHNOLOGY LLC.  
PI Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;  
PI Valentin HE, Venkatesh TV, Venkatesh M;  
XX  
DR WPI: 2002-227151/28.  
XX  
PS P-PDB; AAE19653.  
XX

Claim 3; Page 117-119; 155pp; English.

The invention relates to gcpE nucleic acid molecule, an essential gene of methyl-D-erythritol phosphate (MEP) pathway that encodes rice, Arabidopsis thaliana or Escherichia coli GCPe protein. gcpE is useful for producing a transgenic plant such as Brassica campestris, B. napus, canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard, oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower, or wheat with an increased isoprenoid (tocopherol) compound level. The expression of GCPe protein in organisms increases the level of tocopherol substrate such as isopentenyl diphosphate and dimethylallyl diphosphate biosynthesis. Transgenic organisms overexpressing GCPe protein can nutritionally enhance food and feed sources. Overexpression of GCPe protein in transgenic plant may provide tolerance to stresses e.g., oxidative stress tolerance such as to oxygen or ozone, UV tolerance, etc. gcpE may be used to obtain nucleic acid molecules from the same species, and to obtain nucleic acid homologues. gcpE is also

CC used as or primers. The recombinant vectors are used in plant  
 CC transformation or transfection. gcpb also act as markers capable of  
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).  
 CC gcpb is also used to determine the level or pattern of expression of  
 CC the protein. The present sequence is Escherichia coli gcpb gene.

XX Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other:

Query Match 100.0%; Score 1119; DB 24; Length 1119;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCATTAACCAAGGCTTCAATTAACGTAGAAAATCAACAGTATTATGCGGAAATG 60  
 Db 1 ATGCATTAACCAAGGCTTCAATTAACGTAGAAAATCAACAGTATTATGCGGAAATG 60  
 Qy 61 CCGATTGGGAGATGCTGCTCCATCGCCCTGACATGACCAATACCGGTACGACAGAC 120  
 Db 61 CCGATTGGGAGATGCTGCTCCATCGCCCTGACATGACCAATACCGGTACGACAGAC 120  
 Qy 121 GTGCAAGCAACAGGCTCAATCAATCAAGGCGTGAAGCGGTGCGCTGATTCGTCGT 180  
 Db 121 GTGCAAGCAACAGGCTCAATCAATCAAGGCGTGAAGCGGTGCGCTGATTCGTCGT 180  
 Qy 181 GTATCCGTACCGACGATGACGCGGAGAGCGCTTCAACTCATCAACAGCAGGTTAAC 240  
 Db 181 GTATCCGTACCGACGATGACGCGGAGAGCGCTTCAACTCATCAACAGCAGGTTAAC 240  
 Qy 241 GTGCCCGGTGGGTGACATCCACTTCGATTCGATTCGCGCTGAAATACCGGAAATC 300  
 Db 241 GTGCCCGGTGGGTGACATCCACTTCGATTCGATTCGCGCTGAAATACCGGAAATC 300  
 Qy 301 GCGGTGATTTGCTGCGTATTAAACCTGCAATATCGTAATGAAGAGCGTATTGCGATG 360  
 Db 301 GCGGTGATTTGCTGCGTATTAAACCTGCAATATCGTAATGAAGAGCGTATTGCGATG 360  
 Qy 361 GTGGTTGACTGTGCGCGCATAAAAACATTCGATTCGATTCGCGCTTAAACCGCGATCG 420  
 Db 361 GTGGTTGACTGTGCGCGCATAAAAACATTCGATTCGATTCGCGCTTAAACCGCGATCG 420  
 Qy 421 GTGGAAGAAAGATCTGCAAGAAAGTATGGCAACGACGCGCGAGCGTTCGCGAATCT 480  
 Db 421 GTGGAAGAAAGATCTGCAAGAAAGTATGGCAACGACGCGCGAGCGTTCGCGAATCT 480  
 Qy 481 GGCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 540  
 Db 481 GGCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 540  
 Qy 541 AAAGCGTGTGACGCTTCTCTGCTGTGAGTCTTATCGTTTGTGCAAAACAGATCGAT 600  
 Db 541 AAAGCGTGTGACGCTTCTCTGCTGTGAGTCTTATCGTTTGTGCAAAACAGATCGAT 600  
 Qy 601 CAGCCCTTCATCTGGGGATCAACCGAAGCGGTGGCGGACGCGGCGCAGTAAATCC 660  
 Db 601 CAGCCCTTCATCTGGGGATCAACCGAAGCGGTGGCGGACGCGGCGCAGTAAATCC 660  
 Qy 661 GGCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 720  
 Db 661 GGCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 720  
 Qy 721 GCGGCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 780  
 Db 721 GCGGCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 780  
 Qy 781 CGTTCCGAGGAGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 840  
 Db 781 CGTTCCGAGGAGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 840  
 Qy 841 ATCGGTACGGTTAAACGCGTGAAGCAACGCGTGAAGATATCATCTCGATGAGACGTT 900  
 Db 841 ATCGGTACGGTTAAACGCGTGAAGCAACGCGTGAAGATATCATCTCGATGAGACGTT 900  
 Qy 901 TCGATTATCGGCTGCTGCTGGAATGGCCAGGTGAGGCGCTGCTTCTACACTGCGCGTC 960

Db 901 TCGATTATCGGCTGCTGCTGGAATGGCCAGGTGAGGCGCTGCTTCTACACTGCGCGTC 960  
 Qy 961 ACCGCGGCAACAGAAAAGCGGCTTATGAAGATGGCGTGCAGAAAGACCGTCTGAC 1020  
 Db 961 ACCGCGGCAACAGAAAAGCGGCTTATGAAGATGGCGTGCAGAAAGACCGTCTGAC 1020  
 Qy 1021 AACACGATATGATGACGACGATGAGACGACGATTCGTGCAAGAACGACGATGAC 1080  
 Db 1021 AACACGATATGATGACGACGATGAGACGACGATTCGTGCAAGAACGACGATGAC 1080  
 Qy 1081 GAAGCGGCTGCAATGACGTTGACGAGGTTGAAAATAA 1119  
 Db 1081 GAAGCGGCTGCAATGACGTTGACGAGGTTGAAAATAA 1119

RESULT 5  
 AAT42063

ID AAT42063 standard; DNA; 1830121 BP.

AC AAT42063;

DE 14-SEP-1999 (first entry)

DE Haemophilus influenzae complete genome sequence.

Genome; bacterium; Haemophilus influenzae; computer readable medium;  
 expression modulating fragment; regulation; gene expression; vector;  
 organism; open reading frame; ORF; ds.

OS Haemophilus influenzae.

PN WO9633276-A1.

PD 24-OCT-1996.

PF 22-APR-1996; 96WO-US05320.

PR 07-JUN-1995; 95US-0487429.

PR 21-APR-1995; 95US-0426787.

PR 07-JUN-1995; 95US-0476102.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (UYJO) UNIV JOHNS HOPKINS.

PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

PI WPI; 1996-485782/48.

PT Haemophilus influenzae Rd genome recorded on computer readable

PT medium - useful for identifying commercially important nucleic acid

PT fragments by homology searching

PS Claim 1: Page 77.2-77.1091; 1291pp; English.

CC This sequence represents the complete genome sequence of the bacterium  
 CC Haemophilus influenzae strain Rd. The invention relates to a computer  
 CC readable medium (CRM) having recorded upon it the complete H. influenzae  
 CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide  
 CC sequence at least 99% identical to (I). By providing the full-length  
 CC genomic sequence in a computer readable form, it is possible to identify  
 CC commercially important nucleic acid fragments and expression modulating  
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to  
 CC regulate the expression of a nucleic acid molecule. Vectors and altered  
 CC organisms comprising the predicted ORFs can be used to produce any of the  
 CC polypeptide fragments of the H. influenzae Rd genome.

SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 55.6%; Score 622; DB 17; Length 1830121;  
 Best Local Similarity 74.8%; Pred. No. 3.4e-169;  
 Matches 794; Conservative 0; Mismatches 265; Indels 3; Gaps 1;



QY 755 TCGATATTTTGAATTCGCTGCGTATCCGTCGCGAGGATCATCGCTCCGCCGA 814  
 DB 2615 TCGATATTTTGAATTCGCTGCGTATCCGTCGCGAGGATCATCGCTCCGCCGA 2674  
 QY 815 CCTGTCGCGTACGAGATTTATGTTATCGGTACGCTTAAACGCGCTGAGACAGCGCTGG 874  
 DB 2675 CCTGTCGCGTACGAGATTTATGTTATCGGTACGCTTAAACGCGCTGAGACAGCGCTGG 2734  
 QY 875 AAGATATCATCTCCGATGAGACGTTTGATTCGCTGCGTGGTGAATGCGCCAGGTG 934  
 DB 2735 AAGATATCATCTCCGATGAGACGTTTGATTCGCTGCGTGGTGAATGCGCCAGGTG 2794  
 QY 935 AGGCGTGGTCTTCTACACTCGCGCTACCGCGCGGACACAAAGAGCGGCTCTATGAG 994  
 DB 2795 AGGCGTGGTCTTCTACACTCGCGCTACCGCGCGGACACAAAGAGCGGCTCTATGAG 2854  
 QY 995 ATGGCGTGGCGAAAGACCGCTGAGACAAACATATGATGACCGCTGAGACAGCGCA 1054  
 DB 2855 ATGGCGTGGCGAAAGACCGCTGAGACAAACATATGATGACCGCTGAGACAGCGCA 2914  
 QY 1055 TTCGTGCGAAGCCAGTCAGCTG 1077  
 DB 2915 TTCGTGCGAAGCCAGTCAGCTG 2937

RESULT 7  
 ID ABA92787 standard; DNA: 640681 BP.

AC ABA92787;  
 DT 27-MAR-2002 (first entry)

DE Buchnera sp. genomic DNA SEQ ID NO:1.

KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;  
 KW circular; ds.

OS Buchnera sp.

PN JP2001292771-A.

PD 23-OCT-2001.

PE 07-APR-2000; 2000JP-0107160.

PR 07-APR-2000; 2000JP-0107160.

PA (RIKA) RIKAGAKU KENKYUSHO.

DR WPI; 2002-126043/17.

A genomic DNA of cockroach-symbiotic bacterium

Claim 1: Page 16-230; 237pp; Japanese.

The present invention describes a gene (I) derived from Buchnera sp. containing the DNA (a) or (b), (a) has a fully defined base pair sequence selected from a table of sequences found in the Buchnera sp. genome DNA of ABA92787 given in the specification or is a DNA selected from complementary DNA sequences, and (b) is a DNA which hybridizes with the DNA (a) and encodes a protein. Also described are: (1) a recombinant vector (II) containing (I); (2) a transformant (III) containing (II); (3) a genomic DNA of Buchnera sp. containing the sequence given in ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or (d), (c) is a DNA containing a fully defined sequence given in ABA92788 or ABA92789 and (d) is a plasmid which hybridizes with a DNA; and (5) a method for the preparation of a protein in which (III) is cultured and the expression protein of the objective protein is collected from the resultant culture. The DNA is useful for developing agricultural chemicals for exterminating cockroaches. The present sequence represents the specifically claimed Buchnera sp. genomic DNA sequence, from the

CC present invention.  
 XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;  
 SO Query Match 36.1%; Score 403.6; DB 24; Length 640681;  
 Best Local Similarity 62.2%; Pred. No. 7.2e-106;  
 Matches 653; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

QY 19 ATTCAACGTAAATCAACACGATATTACGTTGGGATGTGCGGATGCGTCT 78  
 DB 314293 ATCATAGAAATAAATTCGATCGTATTTATGTGAAAAGTCGATTTGGCAATTAATCG 314352  
 QY 79 CCGATGCGGTACAGTCCATGACCAATACGGGTACGACAGCGTCGAAAGCAAGCTCAT 138  
 DB 314353 CCAATATGATTCATATCTATGACAAATATCTGCTACTATTAATCTGAAAGCATTAAT 314412  
 QY 139 CAATTCAGGGCGGTGGAACGGGTGGCGCTATATCGTCCGTATCCGTACCCAGCAT 198  
 DB 314413 CAATCTTAGAGTTACAAAGATAGAGTATGTTGCTATTTCTATACCCATTTTA 314472  
 QY 199 GACGCGGAGAGCGGTTCACATCAACAGCAGTTAAAGTCCGCTGGTGGTGC 258  
 DB 314473 AAAGCTGAGAAATCATTTCAAGAAATATAAAGCAACAAATGTTCATGATTCAGAT 314532  
 QY 259 ATCCACTTGAATATCGATTCGCTGTAAGTACGAGTAAGCGGATACGGCGTCAATTTGCTGCT 318  
 DB 314533 ATACATTTTGAATACAGATTTAGCTTTACAGCTATTAATATGTTGTCAGATGTTTAA 314592  
 QY 319 ATTAAACCTGGCAATATGAGTAAAGAGCGTATTCGCAATGCTGATGCTGCGGCG 378  
 DB 314593 ATTAACTCTGGAAATATGGAATAAAGAGATATCAGAAATCATTTCTTACCAANA 314652  
 QY 379 GATTAACATTCGATCCGATCCGATTTAGCGTTAAAGCGCGGATGCGTGAAGAAAGATCGCA 438  
 DB 314653 GATGAATAATTTCCATTTTCGATTTGCTTTAATGCTGATCTTTTGAAGAAAGATATTA 314712  
 QY 439 GAAAGATATGCGGAACGACGCGCGGCTGCTGTAATCTGCCATCGCTCATGTTGAT 498  
 DB 314713 AAAAATATTAATAATCTACATCTCAAGTATGATGATGATGATGATGATGATGAT 314772  
 QY 499 CATCTCGATCCGCTCAACTGTCGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 558  
 DB 314773 TACTTGTATGCTTTAAATTTTAAATTTTAAAGTATGATTTAAAGCGCTGATGATTT 314832  
 QY 559 CTCGCTGTTAGTCTTATGCTGTTGCTGCAAAACAGATCATGATCCGCTGATCGGAG 618  
 DB 314833 TTAGCTATTTGATCATATTCGATGATTTAGGAAAGAAATTAACACACCTTTGATTTGG 314892  
 QY 619 ATCACGGAAGCGGCTGTCGCGGACGCGGCGAGTAAATCCGCAATTTGATTTAGCTG 678  
 DB 314893 ATAACTGAATCCGCTGTTTAAAGTAATGAAACAGTAAATCATCTATGATTTCTTT 314952  
 QY 679 CTGCTGCTGAAGGATGAGGAGGACGCGGCGGCTGATGCGGCGGATCGGCTGCA 738  
 DB 314953 TTATTATTGAAGGATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 315012  
 QY 739 GAGATCAAGTCTGATTTGATTTGAAATGCTGCTGATCCGTCGAGAGGATCAAC 798  
 DB 315013 GAAGTAAAGTATGATTTAGATTTTAAAGTATTTTAAAGTATTTTAAAGTATTTTAA 315072  
 QY 799 TTGATGCGCTGCGGACGCTGTCGCGTCAAGATTTGATTTATGCTGATGCTTAAACCG 858  
 DB 315073 TTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 315132  
 QY 859 CTGAGGCAAGCGCTGAGATATATCATCTCGATGAGAGCTTTGATTTAGGCGGCG 918  
 DB 315133 CTAGAAAAAATCTAGAAATATCTCGACTCCATAGATGATGATTAATTTGCTGCGCT 315192  
 QY 919 GTGATGCGCCAGGAGGAGCGGCTGTTCTACACTGCGGCTGACCGGCGCAACGAAGA 978  
 DB 315193 GTTAATGGAATAGTGAATCTTAATATGCACTTTAGGCTGTAAGCAAGATCATATGAAGA 315252  
 QY 979 AGCGGCTCTATGAAATGAGGCT--GCACAAGACCGCTTGGACACAAACGATATGATC 1035

Db 315253 AGTCCATTATATGAGACGGGTAGACAAAGAAAAATAAAAAGAAATTTATA 315312  
OY 1036 GACCAGCTGGAGCAGCATTCGTGCGAAA 1065  
Db 315313 GAAAAAATGGAAATTTAAAAATTCGAAAAAA 315342

## RESULT 8

AAF28536

ID AAF28536 standard; DNA: 33140 BP.

AAF28536;

04-APR-2001 (first entry)

Genomic fragment #23.

Genomic library; bacteriophage; human upper airway; otitis media; sinusitis;

bronchopulmonary; endocarditis; meningitis; ss.

Moraxella catarrhalis.

MO200078968-A2.

28-DEC-2000.

16-JUN-2000; 2000MO-US16649.

18-JUN-1999; 99US-0140121.

(INCY-) INCYTE GENOMICS INC.

Lagace RE, Patterson C, Berg KL;

WPI; 2001-041427/05.

Genomic library for identifying diagnostic and therapeutic  
compositions, and for identifying virulence factors, regulatory  
elements and drug targets, comprises Moraxella catarrhalis nucleic  
acids -

Claim 1; Page 191-199; 545pp; English.

The present invention relates to a Moraxella catarrhalis genomic library  
comprising of a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28554). The library has a number of uses described in the  
specification e.g. is useful for identifying diagnostic and therapeutic  
compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
aerobic, gram-negative diplococcus, normally found among the bacterial  
flora of human upper airways. M. catarrhalis is known to cause acute,  
localised infections such as otitis media, sinusitis and bronchopulmonary  
infection and life-threatening, systemic diseases including endocarditis  
and meningitis.

Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

Query Match 34.2%; Score 382.6; DB 22; Length 33140;  
Best Local Similarity 59.2%; Pred. No. 2.6e-100;  
Matches 652; Conservative 0; Mismatches 449; Indels 0; Gaps 0;

OY 1 ATGCATTAACCAAGGCTCCATTTCAAGTAGAAATCAACAGCTATTTCAGTTGGGAATGTG 60

Db 21735 ATGTCAATACACAATCGATTAAAGCGCCGCTAACCAAAAAATTCATGTGGCAATGTC 21794

OY 61 CCGATTGGCGATGTGCTCCATCGCCGCTAGACATTCATGACCAATACGCGTACGACAGAC 120

Db 21795 GCCATGGCGGCGCATGCACCATCAGCTGCAAGATATGCAATCAACACCTGTGAT 21854

OY 121 GTGCAAGCAACGCTCAATCAATCAAGGCGCTGCAACGCGCTGTGATTCGTCGCT 180

Db 21855 ATTGATGCAACAGCTGCAACAAATTCAGCGATGCTGAGCAGCGAGTGCGGCTGATGTCGT 21914

OY 181 GTATCCGTACCGAGATGACGCGGACAGACGCTTCAACTCAATCAACAGACAGCTTAAAC 240

Db 21915 GTATGACACCAACAGATGAGAGAGGTGCTGCTTCTGTGATTAATAAAAGAGAGTATCC 21974

OY 241 GTGCGGCTGGTGTGATCATTCACCTTGTGACTATGCAATTCGCTGGAATAGCGGAATAC 300

Db 21975 ATTCCACTCATATGCGGATATTCATTTGATTCACAAAATCCCATTTGACGAGCGGATGTG 22034

OY 301 GCGCTGATGTGTGCGATTAACCGGCAATATGCGTATGTAAGAGCGGATATTCGATG 360

Db 22035 GGTGACACTGCTGCTGATCAATCCAGTATATTTGCAATGTATCAAAAAGTCAAAAGAA 22094

OY 361 GTGTTGACTGTGCGCGCATTAATAACATTCGATCCGATATTCGCTTAAACGCGGATCG 420

Db 22095 GTGTTGACGCGGACGCGCATTAATATGCGGATTCGCTATTCGCTGATTCGATTCGCA 22154

OY 421 CTGGAATAAGATCTGCAAGAAAGTATGCGGACGACGCGGCGCTTGTGGAATCT 480

Db 22155 CTTGAAAGATTTACAAAAAATATCGAAGAACCGACGCTGAGCGGATCTAGAAATCG 22214

OY 481 GCCATGGCTATGTTGATCAATTCGATTCGCTGATTCGATTCGATTCGATTCGATTCG 540

Db 22215 GCGTTACGCCACATTTGATTTAGAAAATCTTATTTAGAAATTTAAATCTCAGTC 22274

OY 541 AAAGCTCTGACGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

Db 22275 AAAGCTCTGACGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22334

OY 601 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

Db 22335 AATTCATTTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22394

OY 661 GCCATTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

Db 22395 GCCATTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22454

OY 721 GCGGCGCATCGCTGCAAGAGATCAAGTGGTTTGAATTTGAATGCTGCTGATC 780

Db 22455 GCGGCGCATCGCTGCAAGAGATCAAGTGGTTTGAATTTGAATGCTGCTGATC 22514

OY 781 CTTTGGGAGGATCAACTTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

Db 22515 CTTTGGGAGGATCAACTTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22574

OY 841 ATCGGTACGCTTAAACGCTGAGCAACGCTGAGATATTCATTCATTCATTCATTCATTC 900

Db 22575 ATCGGTACGCTTAAACGCTGAGCAACGCTGAGATATTCATTCATTCATTCATTCATTC 22634

OY 901 TCGATTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

Db 22635 TCGATTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22694

OY 961 ACCGCGCGCAACAAGAAACGCGCTCTATGAAGATGCTGCTGCTGCTGCTGCTGCTGCT 1020

Db 22695 GTGCGGACGCGCGCAATTAATTTGCTTATTAATTAATTAATTAATTAATTAATTAAT 22754

OY 1021 AACACGATATGATTCGACGCTGAGCAACGCTGATTCGCTGCAAGCACTGCTGAGAC 1080

Db 22755 AACACGATATGATTCGACGCTGAGCAACGCTGATTCGCTGCAAGCACTGCTGAGAC 22814

OY 1081 GAAAGCGCTGGAATTCAGCTT 1101

Db 22815 GAAAGCGCTGGAATTCAGCTT 22835

## RESULT 9

AAS90066

ID AAS90066 standard; cDNA: 1857 BP.

AAS90066;

13-FEB-2002 (first entry)

DE	DNA encoding novel human diagnostic protein #25870.	
XX		
KW	Human: chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
PD		
XX	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(Hkse-) HXSEQ INC.	
PI		
PI	Dermanac RT, Liu C, Tang YT;	
XX		
DR	WPI: 2001-639362/73.	
XX	P-PSDB; ABG25879.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity -	
XX		
PS	Claim 1; SEQ ID NO 25870; 103pp; English.	
XX		
CC	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC	and gene mapping, and in recombinant production of (II). The	
CC	polynucleotides are also used in diagnostics as expressed sequence tags	
CC	for identifying expressed genes. (I) is useful in gene therapy techniques	
CC	to restore normal activity of (II) or to treat disease states involving	
CC	(II). (II) is useful for generating antibodies against it, detecting or	
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	
CC	a food supplement. (II) and its binding partners are useful in medical	
CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
CC	disorders involving aberrant protein expression or biological activity.	
CC	The polypeptide and polynucleotide sequences have applications in	
CC	diagnostics, forensics, gene mapping, identification of mutations	
CC	responsible for genetic disorders or other traits to assess biodiversity	
CC	and to produce other types of data and products dependent on DNA and	
CC	amino acid sequences. AAS64197-AAS94564 represent novel human	
CC	diagnostic coding sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp://wipo.int/pub/published_pcl_sequences.	
XX		
XX	Sequence 1857 BP; 463 A; 472 C; 530 G; 392 T; 0 other:	
XX		
Query Match	31.8%; Score 356; DB 23; Length 1857;	
Best Local Similarity	97.9%; Pred. No. 4.2e-93;	
Matches 466; Conservative	0; Mismatches 0; Indels 10; Gaps 10;	
QY	644 GCGGGGAGATAAATCCGGCATTTGATTAGTCTGCTGCTGTGTAAGGATCGGCGACA 703	
DB	512 GCGGGGAGATAAATCCGGCATTTGATTAGG-CTGCTGCTGTGTAAGGATCGGCGACA 570	
QY	704 CGCTCGCGTATCCGTCGGCGGCCGATCCGCTGGAAGATCAACATCGGTTTCGATATT 763	
DB	571 CGCTCGCGTATCCGTCGGCGGG--CGATCCGCTGGAAGATCAAAAGTCGGTTTCGATATT 629	
QY	764 TGAATTCGTCGGTATCCGTTTCGCGAGGATCAACTTCATCCGCTGCCAGACTGTTCCG 823	
DB	630 TGAATTCGTCG-TATTCGCTTCGCGAGGATCAACTTCATCCGCTGCCAGACTGTTCCG 688	
QY	824 GTACAGATTGATGTTATTCGCTAGCGTTAAAGCGCTGGAGCAACCCCTGGAAGATATCA 883	
DB	689 G--CAGGAATTTATGTTATTCGCTAGCGTTAAAGCGCTGGAGCAACCCCTGG--AGATATCA 746	

OY 884 TCACTCGCATGAGAGCTTTTCGATTTATTCGCGCTGGTGTGAATAGGCCACAGTGGAGCGCTGG 943  
 DB 747 TCACTCGCATGAGAGCTTTTCGATTTATTCGCGCTGGTGTGAATAGGCCACAGTGGAGCGCTGG 805  
 OY 944 TTTTCACACTCGCGGCTCACCGCGGCGGCAACAGAAAAAGCGGCGCTCTATATAGATGGCGCTGC 1003  
 DB 806 TTTTCACACTCGCGGCTCACCGCGGCGGCAACAA-AAAAGCGGCGCTCTATATAGATGG-GTGC 863  
 OY 1004 GCAAGACCCGCTGTGACACAAACAGATATATGATGACACAGCTGGAGACAGCATTTGCTGGCA 1063  
 DB 864 GCAAGACCCGCTGTGACACAA-ATATGATGATGACACAGCTGGAGACAGCATTTGCTGGCA 922  
 OY 1064 AAGCCAGTCAGTGTGACGAGCAAGCGCGCTGCATATGACGTTACAGAGGTTGAAAAATTA 1119  
 DB 923 AAGCCAGTCAGTGTGACGAGCAAGCGCGCTGCATATGACGTTACAGAGGTTGAAAAATTA 977  
 RESULT 10  
 ID ABK74401  
 XX ABK74401 standard; DNA; 1083 BP.  
 AC ABK74401;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Bacillus licheniformis genomic sequence tag (GST) #1692.  
 XX  
 KM Differential gene expression; genomic sequenced tag; GST:  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.  
 XX  
 OS Bacillus licheniformis.  
 XX  
 PN WO200229113-A2.  
 XX  
 XX 11-APR-2002.  
 XX  
 PF 05-OCT-2001; 2001MO-US31437.  
 XX  
 PF 06-OCT-2000; 2000US-0680598.  
 PR 27-MAR-2001; 2001US-279526P.  
 XX  
 PA (NOVO ) NOVOZYMES BIOTECH INC.  
 PA (NOVO ) NOVOZYMES AS.  
 XX  
 XX Berka R, Clausen IG;  
 PI WPI; 2002-416684/44.  
 XX  
 DR  
 XX  
 PT Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic,  
 PT sequenced tag array -  
 XX  
 PS Claim 4; SEQ ID NO 1692; 200pp; English.  
 XX  
 XX The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions,  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array

CC equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 1083 BP; 314 A; 278 C; 265 G; 226 T; 0 other;

Query Match 26.2%; Score 293.4; DB 24; Length 1083;

Best Local Similarity 56.6%; Pred. No. 5.3e-75;

Matches 343; Conservative 0; Mismatches 416; Indels 0; Gaps 0;

22 CAACGTAGAAAATCAACAGGATTTACGTTGGAATGCGGATGGGATGGTGTCTCC 81  
 21 CATGCTCAAAAACGCGTGGCTTAAGTGGACCTTTACAAATAGCGGCAATACGAA 80  
 82 ATGCGCGTACAGTCATGAGCAATACGCGTACGACAGAGTGAAGCAACGCTCAATCA 141  
 81 GTCGTCAATTCACACATGACGACAAAACACATGACGTTGAAGCAACCGTCGCCGAA 140  
 142 ATCAAGGCGCGTGAAGCGTGGATGATGCTGATGCTGATCCGACATGAC 201  
 141 ATCAAGAGCTCGCGGAGAGGATGTCAAATCGTCGCTGCTGATGAACGG 200  
 202 GCGGACAGAGCGTTCAAGGATCAACAGAGGTTAAGTGGCTGCTGATGACATC 261  
 201 GCTGCGACGCGCATTCAGAGATCAAAAAGGATATCCATCCCTGTTGCTGATAT 260  
 262 CACTGCACTATGCGATGCTGCTGAAGTACGGAATACGCGCTGATTTGCTGCTATT 321  
 261 CATTCAACATTAATTTGGGATTAAAGCATCAAGGCGGACGATTAATAATCCGATC 320  
 322 AACCTGCGCATATGCGTAAAGAGAGCGTATTCGATGGTGGTGGCTGCGCGCAT 381  
 321 AATCCGGGTAAATCGCGCGCGGAGAAAGTTTAAACGCTGCTCAACGCAAGAGAA 380  
 382 AAAACATTCGATCGGTATGCGGTTAACCGCGATCGTGAAGAAAGATCTGCAAGAA 441  
 381 AAGGCGATTCGATCGGATGCGGATCAATGACAGGCTCTCTGGAAAAACGATCTTGAG 440  
 442 AAGATGCGCAACCGAGCGGAGCGCTGCTGGAATCTGCCATGCTCATTTGATCAT 501  
 441 AAGTACGCGTATCGGAGGAGGAGCGGATGCGAAAGCCCGCCACCATTAAT 500  
 502 CTGATGCGCTGAATCGATCAAGTCAAGTACGCTGAAGAGCGTGGACGCTTCTC 561  
 501 CTGAGAGATCTCGATTCGAGATATCATGTCACATGAAGCGCTGTGATGTAACCTG 560  
 562 GCTGTTAGCTTATGCTTCTGCGCAAAACAGATCGATCAGCGCTTGATCTGGGATC 621  
 561 GCGATGAGGATGAGTAAAGAGCGGCTTAAGCTTTCGATTCGCTTCAATTAAGCAT 620  
 622 ACCGAGCGGATGCGGAGGCGGAGTAAATCCGCCATGGTTTAAAGTCTGCTG 681  
 621 ACCGATCGGAGAACCTGTTGCGGAGTAAAGCGCGGCTGCTGCGGCGCATC 680  
 682 CTGCTGAAGCATCGGAGCGTGGCTGCTGATCGCTGCGGCGGATCGCTGAAG 741  
 681 CTTCACAAAGATTCGCAATACATTTGCGGATTTCTTAAGCGCGGACCGGCTGAAGAA 740  
 742 ATCAAGTCTGCTTCAATTTGGAATCGCTCGATTCGCTTGGCGAGGATCAATTC 801  
 741 GTAAAAAGTCCGAGGAGCTGAATTTTGGGCTGCTTCAATTCGCGCAACATTTG 800  
 802 ATGCGCTGCGCGACCTGCTGCGTCAAGATTTGATGTTTTCGTACGATTAACCGCTG 861  
 801 ATTCTCGCCGACCTGCGCGGATCGAATGGAATTTGATTCGATTCGCAATGAATTC 860  
 862 GAGCAACGCTGGAAGATATGATCCTGATGAGGATGATTCGATTCGCTGCGCTGCTG 921  
 861 GAAATTCATCTCGCAAAAGAGGCGCGCATCAAAAGTTGCGCTGCTGCTGCTGCTG 920

QY 922 AATGCCAGGTGAGGCGCTGTTTACACTCGCGCTACCGCGGCAACAGAAAG 980  
 DB 921 AACGCTCCGAGAGAGCGCGGAGCGGATTCGAGTCCGCGCGCACGCGGTAATG 979

# RESULT 11

AB070939  
 ID AB070939 standard; DNA: 6157 BP.

AC AB070939;

DT 29-AUG-2002 (first entry)

DE Listeria monocytogenes 4b contig DNA sequence #881.

KW Antibacterial; Listeria; food contamination; mutational analysis;

OS Infection; ds.

XX Listeria monocytogenes 4b.

PN W0200228891-A2.

PD 11-APR-2002.

PF 04-OCT-2001; 2001WO-FR03061.

PR 04-OCT-2000; 2000FR-0012697.

PA (INSP ) INST PASTEUR.

PI (CNRS ) CNRS CENT NAT RECH SCI.

PI Kunst F, Glaser P;

DR WPI; 2002-332479/37.

PT New genomic sequences from Listeria species, useful for detection,

PT treatment and prevention of infection, also related polypeptides,

PT antibodies and modulators

PS Claim 14; SEQ ID 3752; 180pp; French.

XX The present invention relates to nucleic acid sequences

CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of Listeria (e.g. as

CC contaminants in foods, or mutational analysis) and for analysis of

CC gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication

CC and pathogenicity of Listeria (potential therapeutic agents), also for

CC treating infections by Listeria, and are useful as immunogens in

CC anti-Listeria vaccines.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 6157 BP; 1731 A; 1065 C; 1209 G; 2150 T; 2 other;

Query Match 22.8%; Score 254.6; DB 24; Length 6157;

Best Local Similarity 53.7%; Pred. No. 2.2e-63;

Matches 527; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

QY 18 AATCAACGTAGAAAATCAACACGATTTTACGTTGGGAATGCGGATGGCGATGTC 77  
 DB 3116 AATCTTCGCGAAACACTCGGCGCATCCAAAGTAGTAATTTAATTTAGTGTGTA 3175  
 QY 78 TCCCATCGCGGTACGATGCAATGCAATAGCGGTACGACAGAGCTGAAGCAACGTTCA 137  
 DB 3176 GGAATTAATCTATCAAAAGATGACTACTACAAAGACATGATGTGAAGCAACGTA 3235  
 QY 138 TCAATCAAGGCGCTGAGACGCTTGGCGCTGATTCGCTGATTCGCTGATTCGAT 197  
 DB 3236 GGAATTAACCGATTTAGAAAGAGGTGCTGATGATTTGCGAGTGTCTTCTGATGA 3295





Db 492 AATTCGAGAGATTACAGATTATGATATCATTCATTGTTCTTGAAGGCTTCGTATGAA 551  
 QY 558 CCTGCTGTTGAGTCTTATGTTGTTGCTGGCAAAACAGATCGATCGCCGTTCATCTGGG 617  
 Db 552 TTATGACATTAAGCTTATGATTAAGTACAGCGCATTTATATCTCTGCACTCCGG 611  
 QY 618 GATACACGAGCCGCTGGTGGCCAGCGGGGCACTAAATCCGCAATTCGTTAGTCTT 677  
 Db 612 AATTACAGAAATCTGCTACAGAAATTTGCTGGAGAAATTAAGAATCTCTGTTAGAGAC 671  
 QY 678 GCTCTGCTGCAAGGAGCATGGCCAGCTGCGGTATTCGCTGGCCGCGCATCCGGTGA 737  
 Db 672 GATACACTAGTTTGGGCACTTGGAAATACATTAACGATTAATCTTGTGCTGATCTGGA 731  
 QY 738 AGACATCAAGTGGTTCGATTTTGAATCCGTCGATCCGTCGCGAGGATCA 797  
 Db 732 AGAAATTAAGTGGCCGCGGAGTTTAAATATTTGTCCTTCCGATCCGCAAT 791  
 QY 798 CTTCATGCGCTGCGCCGACGCTTGGCTGCAAGAAATTTGATGTTATCGGTACGTTAACG 857  
 Db 792 GCTTATCTCTGCTGCTGCTGCGGTGCAATAGAGATTGATTAATGCTATCCGCTAATGA 851  
 QY 858 GCTGAGCAACGCTGAGAAATATCATCTCGATGAGCTTTCGATTAATCGGCTGCT 917  
 Db 852 AGTGGAAATTAACATACGAAATGATGAAGTTCCGATTAAGTACCGCTGCTGCTGCTGC 911  
 QY 918 GGTGATGCGCCAGGTAGGCTGCTTCTTACACTGCGGCTCACCGCGGCAACAA 977  
 Db 912 GGTCAACGCGCTGAGAAATCTGCGGCAAGCGATATGCAATTCGTTCAACGAGGA 971  
 QY 978 AAGGCGCTCTATGAGATGG 998  
 Db 972 AGGCTTCTTTTGAACATGG 992

## RESULT 13

ID AB069993 standard; DNA; 2107 BP.  
 AC AB069993;

DT 29-AUG-2002 (first entry)  
 DE Listeria monocytogenes EHEC DNA sequence #205.  
 KW Antibacterial; Listeria; food contamination; mutational analysis;  
 OS Listeria monocytogenes EHEC.  
 PN MO200228891-A2.  
 PD 11-APR-2002.  
 PF 04-OCT-2001; 2001MO-FR03061.  
 PR 04-OCT-2000; 2000FR-0012697.  
 PA (INSP) INST PASTEUR.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 PI Kunst F, Glaser P.  
 DR WPI; 2002-332479/37.  
 PT New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators.  
 PS Claim 16; SEQ ID 2806; 180pp; French.  
 CC The present invention relates to nucleic acid sequences  
 CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pcr-sequences.  
 XX

Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Query Match 22.5%; Score 251.4; DB 24; Length 1107;

Best Local Similarity 53.5%; Pred. No. 8.8e-63; Matches 525; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

QY 18 AATTCACGTAAGAAATCAACACGATATTACGTTGGAAATGTGCGCATGCTGC 77  
 Db 12 AATATTTCGGGAAACACCTGCGCCAGTCCAAAGGGTAATTAATTAATGTTGTTAGTGA 71  
 QY 78 TCCCATGCGCGTACAGTCAATGACCAATACCGGTACAGACAGCTGAGCAACGATCA 137  
 Db 72 GGAATTAACATCAAGCATGACTACTACAAGACACATGATGTGCAACCAACAGTACG 131  
 QY 138 TCAATCAAGGCGCTGAGACGCTTGGCGCTGATATGCTCCGTATCCGTACCGACAT 197  
 Db 132 AGAAATTCACCGATTAAGAAAGAGCTGTGTCAGATATGTCGAGATTCCTGCTCATGA 191  
 QY 198 GAGCGCGGCAAGAGCGTTCAACATCAACACAGAGTTACGTGCGCTGCTGCTGCA 257  
 Db 192 AGCTGACGAGAAATGCTTGTGTGCGCATCAAGAAATAATCAATATTCGCTTGTGAGA 251  
 QY 258 CATCACTCTGACATGCAATGCGCTGAAAGTAAAGCAATACGCGCTCATTTGCTGCG 317  
 Db 252 TATTCATTTGATTAACGATAGCACTTAAGCATATGATGACGTTTACCAAAATTCG 311  
 QY 318 TATTAACCTTGGCAATATGCGTAATGAAGCGTATTCGATGCTGCTGCTGCGCG 377  
 Db 312 GATTAATCTCGTAACTGATGCTGCTGCGGTGAGGAGTGAAGATGATTAATCTGCTAA 371  
 QY 378 CGATTAACATTCGCAATGCGTATTTGCGTTAACCGCGGATTCCTGGAAAAATCTGCA 437  
 Db 372 AGCAAAAAATATTCATCCATTCGTAATGCGGTAAATGCTGAGTTTGAAGAAATTAAT 431  
 QY 438 AGAAAGTATGCGGAACCGACGCGGAGCGGTGCTGGAATTCGCATGCGCTATGTGA 497  
 Db 432 TCAAAATATGTTACCTTACCTCTGACGGAATGTTGAAGATGCACTGCCATATTA 491  
 QY 498 TCATCTGATGCGCTGAACTTCGATCAATCAAGTCAAGCGTGAAGCGTGAAGCTTT 557  
 Db 492 AATTCGCAAGATTAGATTTTATGATATCATTTGTTTGAAGCTTCATGATGA 551  
 QY 558 CCTGCGTGTGATCTTATGCTTTGCTGCAAAACAGATGATCACCCGTTGATCTGG 617  
 Db 552 TTATGCAATTAAGCTTAATGAATTAAGCTTAATTAATTAATTAATTAATTAATTA 611  
 QY 618 GATCACCAGAACGCGTGGTGGCGGACGCGGAGTAAATCCGATTTGTTAGTCT 677  
 Db 612 AATTACAGAAATCTGTAACAAATTTCTGAGGAAATTAAGATGCTGCTGTTAGAGC 671  
 QY 678 GCTGCTGTGAAGGATCGGCGACACGCTGCGGTATGCTGCGCGGCGGATCCGATCGA 737  
 Db 672 GATACACTAGTTTGGCATGGAATACATTAACAGATATCTTGAAGTATCTCTGGA 731  
 QY 738 AGAGATCAAGTGGTGTGATTTTGAATGCTGCTATCCGTTCCGAGAGATCA 797  
 Db 732 AGAAATTAAGTGGCCGCGGAGTTTAAATCAATTTGCTTCTCGAATGCCCAAT 791  
 QY 798 CTTCATGCGCTGCGCCGACCTGTCGCGTACGAAATTTGATGATGATGATGATGATGATG 857  
 Db 792 GCTTATCTCTGCGCTACCTTGGCGGTGAATAGAGATTTGATTAATTCGATGCTAATGA 851

QY 858 GCTGAGCAACGCTTGAAGATATCATCTCCGATGACGTTTGGATTATCGCGT 917  
DB 852 AGTGGAAATTAATACAGCAAGATGAGCTGCTTAAAGTAGCCGCTTGGCTGCG 911  
QY 918 GGTGATGCGCCAGGTGAGCGGTGTTCTACACGCGGCTACCGCGGCAACAAGA 977  
DB 912 GGTCAACGCGCTTGGAGAGCTGCGCAAGCCATATCGGAATGCTGTTCAACGAGA 971  
QY 978 AAGCGGCTCTATGAGATG 998  
DB 972 AGGCTTCTTTTATGACATG 992

RESULT 14  
AAS59573/C  
ID AAS59573 standard; DNA; 35829 BP.  
XX AAS59573;  
XX DT 13-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein encoding DNA #68.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX KW dermatological; osteopathic; neuroprotectant; ds.  
XX OS Propionibacterium acnes.  
XX PN WO200181581-A2.  
XX PD 01-NOV-2001.  
XX PE 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX PI L'Alouneuve J, Zhang Y, Jen S, Carter D;  
XX DR WPI: 2001-616774/71.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX PT vaccinating against and diagnosing infections, especially useful for  
XX PT treating acne vulgaris -  
XX PS Claim 1; SEQ ID NO 68; 1069pp; English.  
XX CC Sequences AAS59506-AAS59804 represent DNA molecules encoding  
XX CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
XX CC associated DNA sequences are used in the treatment, prevention and  
XX CC diagnosis of medical conditions caused by P. acnes. The disorders include  
XX CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
XX CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
XX CC in infections of bone, joints and the central nervous system, however it  
XX CC is particularly involved in the inflammatory lesions associated with acne  
XX CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
XX CC patient comprises contacting a sample with a binding agent that binds to  
XX CC the proteins of the invention and determining the amount of bound protein  
XX CC in the sample. The polypeptides may be used as antigens in the production  
XX CC of antibodies specific for P. acnes proteins. These antibodies can be  
XX CC used to downregulate expression and activity of P. acnes polypeptides and  
XX CC therefore treat P. acnes infections. The antibodies may also be used as  
XX CC diagnostic agents for determining P. acnes presence, for example, by  
XX CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
XX CC polypeptides shown in AAS5925-AAS5930.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 35829 BP; 7352 A; 11611 C; 10362 G; 6498 T; 6 other:  
SQ

Query Match 21.8%; Score 244; DB 23; Length 35829;  
Best Local Similarity 53.6%; Pred. No. 5,5e-60;  
Matches 530; Conservative 0; Mismatches 455; Indels 3; Gaps 1;

QY 22 CAACGTAGAAATCAACACAGTATTACGTGGGAATGTGCGATTGGCGATGCTCC 81  
DB 26378 CCAGCTCGTAACAGCCAGACAGTCAAGGTGGAGATGCTGTGTGGTGGCCAGCC 26319  
QY 82 ATCGCGTACAGTCCATGACCAATACCGGTACAGACAGTGGAGCAACGGTCAATCA 141  
DB 26318 ATTTAGTGCATCCATGACCAACCAAGACTCAGACGTTGTGCGACCTGCAAG 26259  
QY 142 ATCAAGCGCGTGAAGAGCGGTGGCGTATATCGTCCGTATCCGATCCGATGAG 201  
DB 26258 ATCGCGCGCTTGACCGCTGACAGGTGTGACATCGTACGCGTGCATGCGCAAC 26199  
QY 202 GCGGCAAGCGGTCAAACTCATCAACAGCAAGTTAACGTGCGCTGTGCTACATC 261  
DB 26198 GACGCGAGGTGCTACCGATCATCGCAAGCGTCCAGATCCGTGATTTGCGATATC 26139  
QY 262 CACTTCGACTATCGCATTTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 321  
DB 26138 CACTTCGACGCGAAGTATGATGCTTCCAGAGATGAGGCTGTGCGCGCGTGC 26079  
QY 322 AACCTGGCATATTCGGTAA--TGAGAGCGTATTCGATGCTGCTGCGCG 378  
DB 26078 AACCCGGAAATATCCGAGTTCGACGACAGATGATGATTTGACAGCTCCAC 26019  
QY 379 GATAAACATTTCCGATCGATGCTTAACGCGGATGCGGAGAAAGATCTCGAA 438  
DB 26018 GAGCAGGTAGAGATCGATCGATCGCTGATGCTGCTGCTGCAAAAGTGTGCTT 25959  
QY 439 GAAAGTATGCGCAACGACGCGCGAGCGTGTGATGATGCTGATGCTGATGAT 498  
DB 25958 GACAAATAGGAGCGCGCGAGCGGAGGATGAGTGTGCGGCTGTGGAGGCTGAC 25899  
QY 499 CATTCGATCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 558  
DB 25898 CTCTTTGAGCAATACGATTCGCGGATTTCAAAATTCGAGAAACACACACCGCTT 25839  
QY 559 CTCGCTGTGAGTCTTATGCTTGTGCGCAAAACGATGATGATGATGATGATG 618  
DB 25838 GTCAATGATCGGTGCTATGACACGCTGCGCGCAAAATGATTTATCCCTTCACT 25779  
QY 619 ATCAACGAGCGGTGTGCGCGAGCGGCGAGTAAATCGGCAATGCTTATGCTG 678  
DB 25778 GTTACTGAGGCTGTGCGCGCTTCCAAAGGACATCAAGTGTGCGGCTGCGG 25719  
QY 679 CTGCTGTGAGGATGATGCGGAGACAGCTGCGGCTATGCTGCGGCGGATTCG 738  
DB 25718 CTCTTGTGCGGATGATGCGGAGACATGATGCTGCTGCTGCTGATGCTGAG 25659  
QY 739 GAGATCAAGTGTGCTTATGATTTTGAATGCTGCTGATGCTGCGGAGGATCAAC 798  
DB 25658 GAGTCAAGTGTGCTTATGATGATGCTGCTGATGCTGCTGCTGCTGCTGCTG 25599  
QY 799 TTCAATGCTGCGCGAGCTGCTGCGGCTGAGATTTGATGATGATGATGATG 858  
DB 25598 ATCTGCTGCTGCGCGCTTGTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTG 25539  
QY 859 CTGAGCAAGCGCTGGAAGATATCATACCTGCTGAGATGCTGCTGCTGCTGCTG 918  
DB 25538 GTTACTGATGCTGCTGCTGGAAGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 25479  
QY 919 GTGATGCGCGGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978  
DB 25478 GTCAATGCGGCTGGAAGAGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTG 25419  
QY 979 AGGCGCTCTATGAGATGAGGCTGCGCA 1006

Db 25418 GGCAGATCTTCAAGCATGGCGAGGTCA 25391  
|| | | | | | | | | | |  
RESULT 15  
AAH52056  
ID AAH52056 standard; DNA: 1164 BP.  
XX  
AC AAH52056:  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 110.  
XX  
KM Drug target; growth; organism viability; characterisation; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN MO20013517-A1.  
XX  
PD 17-MAY-2001.  
XX  
PF 13-NOV-2000; 2000MO-US31152.  
XX  
PR 12-NOV-1999; 99US-0165086.  
PR 12-NOV-1999; 99US-0165124.  
PR 01-FEB-2000; 2000US-0179531.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Eisenberg D, Rotstein SH, Marcotte EM;  
XX  
DR WPI: 2001-329193/34.  
XX  
P-PSDB; AAG81205.  
XX  
PT Identifying nucleotide or polypeptide sequence for use as drug target,  
PT involves providing algorithm that analyzes a functional relationship  
PT between nucleotide or polypeptide sequences, and comparing the  
XX sequences -  
PS Disclosure: Page 124-125; 207pp; English.  
XX  
XX This invention relates to a method for identifying a nucleotide or  
XX polypeptide sequence that may be a drug target, or essential for growth  
XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
XX represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium  
XX tuberculosis proteins which are potential drug targets. The DNA and  
XX protein sequences are used to illustrate the method of the invention. The  
XX method involves providing an unknown nucleotide or polypeptide sequences,  
XX and comparing it to a number of sequences along with at least one  
XX algorithm capable of analyzing a functional relationship between  
XX nucleotide and polypeptide sequences. The method is useful for  
XX characterizing the function of nucleic acids and polypeptides that may be  
XX useful as a target for a drug or essential for the growth or viability of  
XX an organism.  
XX  
XX Sequence 1164 BP; 203 A; 357 C; 402 G; 202 T; 0 other:  
SQ  
Query Match 21.8%; Score 243.8; DB 22; Length 1164;  
Best Local Similarity 53.5%; Pred. No. 1.5e-60;  
Matches 534; Conservative 0; Mismatches 462; Indels 3; Gaps 1;  
QY 49 GTTGGGAATGTGCGATTGGCATGTGCTCCATCGCCGTACAGTCCATGACCAATACG 108  
|| | | | | | | | | | |  
Db 76 GTCCGCAACGTGCGCGTGGCGATGACCATCGGTCTGTCGCAATGATGTGACACC 135  
|| | | | | | | | | | |  
QY 109 CGTACGACAGCGTTCGAAGCAAGCTCATCAATCAAGGCGTGGAAAGCGTTGGCGCT 168  
|| | | | | | | | | | |  
Db 136 AAAACCCACGACGTCACTGACATTCGACAAATCGCCGAGTACCCGCGGATGC 195  
|| | | | | | | | | | |  
QY 169 GATATCGTCCGTATCGTACGACAGTGGACGCGGAGAGGCTTCAAACTCATCAAA 228  
|| | | | | | | | | | |  
Db 196 GACATCGTGGGCTGCTGCGCCGCGCAGAGAGGACGCGGCTGGCGGAGATGCC 255  
|| | | | | | | | | | |

QY 229 CAGCAGTTAAGCTGCCGCTGTGAGTCAATCCACTTGCATTCGACTATTCGCTGAAA 288  
|| | | | | | | | | | |  
Db 256 CGGCACAGCAGATATCCCGAGTGCAGCATATATTCACGCGGCTACATATTCGCC 315  
|| | | | | | | | | | |  
QY 289 GTAGCGAATACGCGCTGATTTGCTGCTATTAACCTTGGCAATATC--GGTATGAA 345  
|| | | | | | | | | | |  
Db 316 GCCATCGACGCTGATGTGCGCGGCTGACATCCGCGCAACATCAAGAGATTGAC 375  
|| | | | | | | | | | |  
QY 346 GAGCGTATTGCGATGCTGCTGCTGCTGCGCGGATATAAAGATTCGATTCGATTCG 405  
|| | | | | | | | | | |  
Db 376 GCGCGGCTGGTGTAGTGTGCGCAAGGCGGCGGATCCGATCCGATTCGATTCG 435  
|| | | | | | | | | | |  
QY 406 GTTAACGCGGATGCTGCAAAAAGATCTGCAAGAAAGATATGCGCAACCGCGCGAG 465  
|| | | | | | | | | | |  
Db 436 GTCAACGCGGCTTCCGTGCAACACGTTATGAGAGATATGCAAAACCGCGCGAG 495  
|| | | | | | | | | | |  
QY 466 GCGTGTGTAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525  
|| | | | | | | | | | |  
Db 496 GCGCTGTGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 555  
|| | | | | | | | | | |  
QY 526 TTCAAAGTACGCTGGAAGGCTGCAAGCTTCCCTGCTGCTGCTGCTGCTGCTGCTG 585  
|| | | | | | | | | | |  
Db 556 ATCAAGATCAGCTGCAAGCAACGACGACGACGACGACGACGACGACGACGACG 615  
|| | | | | | | | | | |  
QY 586 GCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 645  
|| | | | | | | | | | |  
Db 616 GTCACAGGCTGCAAGCAACGACGACGACGACGACGACGACGACGACGACG 675  
|| | | | | | | | | | |  
QY 646 GGGGCAATTAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 705  
|| | | | | | | | | | |  
Db 676 GGCACATCAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 735  
|| | | | | | | | | | |  
QY 706 CTGCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765  
|| | | | | | | | | | |  
Db 736 ATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 795  
|| | | | | | | | | | |  
QY 766 AATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825  
|| | | | | | | | | | |  
Db 796 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 855  
|| | | | | | | | | | |  
QY 826 CAGGAATTTGATGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885  
|| | | | | | | | | | |  
Db 856 GCGCAATGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 915  
|| | | | | | | | | | |  
QY 886 ACTCGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945  
|| | | | | | | | | | |  
Db 916 GTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 975  
|| | | | | | | | | | |  
QY 946 TCTACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005  
|| | | | | | | | | | |  
Db 976 GCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1035  
|| | | | | | | | | | |  
QY 1006 AAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044  
|| | | | | | | | | | |  
Db 1036 ATCAAGACCGTCCGCAAGCAGATGCTGAGACGCTG 1074  
|| | | | | | | | | | |

Search completed: May 10, 2003, 15:48:52  
Job time : 2434 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 15:05:07 ; Search time 1889 Seconds  
(without alignments)  
9593.833 Million cell updates/sec

Title: US-09-921-992-3  
1119  
Perfect score: 1 atgcataaccagctcccat.....ttcacagcttgaataaa 1119  
Sequence:

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.8	7.3	351	14	BU004160 OGG37N12.
2	81.6	7.3	720	17	BH235010 PH.05.x.S
3	67.8	6.1	448	10	AV644003 AV644003
4	67.4	6.0	482	10	AV642898 AV642898
5	67.4	6.0	502	10	AV626792 AV626792
6	67.4	6.0	512	10	AV644476 AV644476

7	67.4	6.0	531	10	AV628026
8	67.4 <th>6.0</th> <td>531</td> <td>10</td> <td>AV644087</td>	6.0	531	10	AV644087
9	67.4 <td>6.0</td> <td>545</td> <td>10</td> <td>AV626844</td>	6.0	545	10	AV626844
10	65.8	5.9	467	10	AV629427
11	62	5.5	377	10	AV641368
12	56.4	5.0	380	9	A1770308
13	54.2	4.8	303	10	AV640429
14	53.8	4.8	597	10	AM282425
15	53.6	4.8	728	12	BG525868
16	53.4	4.8	658	17	A2570993
17	52.6	4.7	640	10	AW738440
18	52.6	4.7	680	12	BG830681
19	51.8	4.6	731	12	BG124857
20	51.4	4.6	492	13	BM323500
21	51	4.6	579	13	B1935552
22	51	4.6	700	13	B1933447
23	50.4	4.5	173	12	BF962770
24	49.8	4.5	576	14	BQ975044
25	49.4	4.4	532	12	BE924238
26	49.4	4.4	645	13	B1934498
27	48.4	4.3	608	13	AV917069
28	48.4	4.3	558	13	B1465443
29	48.4	4.3	616	10	AV934187
30	48.4	4.3	644	13	B1464122
31	48.4	4.3	651	9	A0252386
32	48.4	4.3	672	10	AV916407
33	47.6	4.3	585	10	BE354448
34	47.6	4.3	679	13	B1931456
35	47.4	4.2	534	9	A0252403
36	47.2	4.2	601	10	AM042702
37	46.8	4.2	648	13	B1479404
38	46.6	4.2	670	10	AV821766
39	46.4	4.1	498	10	AM576226
40	45.4	4.1	422	13	B1478402
41	45.4	4.1	528	13	B1482368
42	44.2	3.9	610	13	B1481080
43	44	3.9	465	9	A1482869
44	44	3.9	588	10	AW738232
45	44	3.9	625	10	AW944795

## ALIGNMENTS

RESULT 1  
BU004160/c  
LOCUS  
DEFINITION OGG37N12.yg.ab1 OG\_EFGH1 lettuce serriola Lactuca sativa cDNA clone  
ACCESSION BU004160  
VERSION BU004160.1 GI:22438555  
KEYWORDS  
SOURCE  
ORGANISM Lactuca sativa.  
Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
Lactuca.  
REFERENCE  
AUTHORS 1 (bases 1 to 351)  
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,  
P., Koltman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
Church, S., Jackson, L. and Bradford, K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
http://compogenomics.ucdavis.edu/  
JOURNAL Unpublished (2002)  
COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]







OY	50	TTGGGAATGTGCCGATTGGCGAATGGTCCTCCACATGCCTTACAGTCAATGACAATACC	109
Db	107	TGGGCAAGGTCCGGTGGGGAGCAGCACACCAGCATGCCCTTCAGACCAATGACCACACGG	166
OY	110	GTAACGACACAGCTCGAAGDAGCGGTCAATCAATAATCAAGSCGCTGGAAGCGTTGGCGCTg	169
Db	167	ACACCGCGAATGTGCGAGCTGACCGCTGGACAGTAGTAAGAATGGCGGAGCGGCGCGG	226
OY	170	AATATCGTCGTGATCCGTAACGAGATGAGCGCGGCAAGCGTCA-----	217
Db	227	ACATGCTCCGCATCAGAGTGGCAGGCGCAAGAGAGCTGAGCGTCAATGAAGATTGCGG	286
OY	218	---AACTCATCAAACAGCAGGTTAACGTGCGCGCTGGTGCATGACATCACTTCGACTATC	274
Db	287	ACGACGCTGTTCAAGGACAGCATGACAGCTGCCCTGGTGGCCGACATCCACTTCACGCCA	346
OY	275	GCATTGCGCTGAAAAGTAGCGGAATACGGCGTGGATTGTTCGGTATTAACCTGGCAATA	334
Db	347	CGGTGGCCCATGATGTGTGGCG--ACGCTTCGAGAAGATCGTGTGAACCCCGCACT	403
OY	335	TGGGTATG 343	
Db	404	TGGGTGAGC 412	
RESULT 6			
AV644476		512 bp mRNA linear EST 15-DEC-2000	
LOCUS	AV644476 Chlamydomonas reinhardtii 5k CO2 Chlamydomonas reinhardtii		
DEFINITION	CDNA clone HCL089B10_r 5', mRNA sequence.		
ACCESSION	AV644476		
VERSION	AV644476.1 GI:10787804		
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii.		
ORGANISM	Chlamydomonas reinhardtii.		
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
AUTHORS	Asanizu,E., Muraoka,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.		
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii		
JOURNAL MEDLINE	DNA Res. 7 (5), 305-307 (2000)		
COMMENT	20539644.		
FEATURES	Contact: Erika Asanizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asanize@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/. Location/Qualifiers		
source	1..512		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="HCL089B10_r"		
	/note="Vector: plasmidscript SK-; Site_1: EcoRI; Site_2:		
	XhoI; The cDNA library was constructed from cells cultured		
	in a medium with bubbling air containing 5% carbon		
	dioxide"		
BASE COUNT	118 a 159 g 158 g 77 t		
ORIGIN			
Query Match	6.0%; Score 67.4; DB 10; Length 512;		
Best Local Similarity	56.0%; Pred. No. 1.7e+09;		
Matches 175; Conservative	0; Mismatches 116; Indels 18; Gaps 2;		
OY	50	TTGGGAATGTGCCGATTGGCGAATGGTCCTCCACATGCCTTACAGTCAATGACAATACC	109
Db	96	TGGGCAAGGTCCGGTGGGGAGCAGCACACCAGCATGCCCTTCAGACCAATGACCACACGG	155
OY	110	GTAACGACACAGCTCGAAGDAGCGGTCAATCAATAATCAAGSCGCTGGAAGCGTTGGCGCTg	169

DB	156	ACACCCGGCAATGCGAGCTGACCGCTGAGACCGTGTGAAGAAGTGGCCGAGCGCGGGCGCG	215
QY	170	ATATGTCCTCGTGTATTCCTGATCCGACGATGGACGCGGACAGCGTTCA-----	217
Db	216	ACATGCTCCGCAATCACAGTGCAGCGGCAAGAGAGAGCTGACGTGCAATGAAGATTTCGGC	275
QY	218	---AATCATCAAAAGCAGAGGTTAACTGTCGCTGGTGGCTGACATTCACACTGCACTATC	274
Db	276	AGCAGCTTCAAGGCAAGTACGACGTCCTGGTGGCGACATTCACATTCACAGCCCA	335
QY	275	GCATGCGCTGGAAGATGACGGGATACGCGGTGATTCGTGCTATTAACCTGCAATA	334
Db	336	CGGTGGCCATGATGTGTGGCG---ACGCTTTCAGAAAGATTCGTGTGAACCCGGCACT	392
QY	335	TCGCTAATG 343	
Db	393	TCGCTGACG 401	
RESULT 7	AV628026	531 bp	mRNA
LOCUS	AV628026	Chlamydomonas reinhardtii	5% to 0.04% CO2 Chlamydomonas
DEFINITION	reinhardtii cDNA clone LC1035f07_r 5', mRNA sequence.		
ACCESSION	AV628026	GI:10790660	
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii.		
ORGANISM	Chlamydomonas reinhardtii		
REFERENCE	1 (bases 1 to 531)		
AUTHORS	Asamizu, E., Mura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohnaka, K., Nakamura, Y., and Yabata, S.		
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii		
JOURNAL	DNA Res. 7 (5), 305-307 (2000)		
MEDLINE	20539644		
COMMENT	Contact: Erika Asamizu		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
FEATURES	Location/Qualifiers		
Source	1..531		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="LC1035f07_r"		
	/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"		
BASE COUNT	119 a 166 c 165 g 81 t		
ORIGIN			
Query Match	6.0%; Score 67.4; DB 10; Length 531;		
Best Local Similarity	56.6%; Pred. No. 1.8e-09;		
Matches 175; Conservative	0; Mismatches 116; Indels 18; Gaps 2;		
QY	50	TTGGGAATGCGCGATGCGATGGTGTCCCATCGCGCTGACATGCAATACGC	109
Db	106	TCGGCAAGTCCGGTGGGCGACGACGCCATTCGACATGACCAACGCG	165
QY	110	GTACGACAGACGTCGACGACGATCAATCAATCAAGCGCTGGAACGCTGGCGCTG	169
Db	166	ACACCCGCAATGCTGACGTCGACCGCTGACCGCTGGAAGATGTCGCGACGCGCGCG	225
QY	170	ATATGCTCCGCTGATTCCTGATCCGACGATGGACGCGGCAAGAGCGTTCA-----	217
Db	226	ACATGCTCCGCAATCACAGTGCAGGCGCAAGAGAGGCTGACGTGCAATGAATTCGGC	285

[illegible]

Db	337	CGGTGGCGATGATGTGGCGG---ACGGCTTCGAGAAATATCGGTGAACCCCGGCAACT	393
QY	335	TCGGTATG 343	
Db	394	TCGCTGACG 402	
RESULT	9		
LOCUS	AV626844		
DEFINITION	AV626844 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas.		
ACCESSION	AV626844		
VERSION	AV626844.1		
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii.		
ORGANISM	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
REFERENCE	Chlamydomonadaceae; Chlamydomonas.		
AUTHORS	1 (bases 1 to 545)		
TITLE	Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,		
JOURNAL	Nakamura,Y. and Tabata,S.		
MEDLINE	Generation of expressed sequence tags from low-CO2 and high-CO2		
COMMENT	adapted cells of Chlamydomonas reinhardtii		
FEATURES	DNA Res. 7 (5), 305-307 (2000)		
Source	Contact: Erika Asamizu		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
	Location/Qualifiers		
	1..545		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="lcl016e04_r"		
	/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"		
	/note="Vector: p Bluescript II SK-; Site 1: EcoRI; Site 2:		
	XhoI; The cDNA library was constructed from cells cultured		
	in a carbon stress acclimatized condition in which carbon		
	dioxide concentration in the bubbling gas was changed from		
	5% to 0.04%"		
BASE COUNT	121 a 177 c 166 g 81 t		
ORIGIN			
Query Match	6.0%; Score 67.4; DB 10; Length 545;		
Best local similarity	56.6%; Pred. No. 1.8e-09;		
Matches 175; Conservative	0; Mismatches 116; Indels 18; Gaps 2.		
QY	50	TTGGGATGATGCGGATGTGGTGTGCTCCATGCGCGGTACAGTTCATGACCAATACGC	109
Db	90	TCGGCAAGGTGGCGGTGGGACACGACACCCGATCGCCCTTCAAGACATGACCAACACG	149
QY	110	GTACACAGACGTGCAAGCAACGGTCATCAATTAAGAGCGCTGGAACGCCCTGGCGCTG	169
Db	150	ACACCCGGAATGTGAGGTGACGTGACGACCAAGTAAGAGTGGCCGCGACGCGCGCGG	209
QY	170	ATATGCGCGTATCCGATACGACGAGATGGAGCGGCGAGAAACGTTCA-----	217
Db	210	ACATGTCGCGATCCAGTGCAGGCGCAAGAGAGGCTGAGGCGTGCATGAAGATTCCGG	269
QY	218	---AACTCATCAACAGCAGTTAAGCTGCGCGCTGGTGGCTGACATCCACTTCGACTATC	274
Db	270	AGCACTCTTCAAGAGCAAGTACGACAGTGGCCCTGGTGGCGCAATTCACCTTCAGGCCA	329
QY	275	GCATTGGCGCTGAAGTACCGGATATCGGCGTTCGATTTCGCTTTAAACCTTGGCAATA	334
Db	330	CGGTGGCGATGTGGCGG---ACGGCTTCGAGAAATATCGGTGAACCCCGGCAACT	386
QY	335	TCGGTATG 343	

DB	387	TCGCTGACG	395
RESULT 10			
LOCUS	AV629427	467 bp	mRNA linear EST 15-DEC-2000
DEFINITION	AV629427 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas		
ACCESSION	reinhardtii.cdna clone LCLO58F01_r 5', mRNA sequence.		
VERSION	AV629427		
KEYWORDS	AV629427.1 GI:10792061		
SOURCE	EST.		
ORGANISM	Chlamydomonas reinhardtii.		
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
AUTHORS	1 (bases 1 to 467)		
TITLE	Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S.		
JOURNAL	Generation of expressed sequence tags from low-CO2 and high-CO2		
MEDLINE	adapted cells of Chlamydomonas reinhardtii		
COMMENT	DNA Res. 7 (5), 305-307 (2000)		
FEATURES	20539644		
SOURCE	Contact: Erika Asamizu		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
	Location/Qualifiers		
	1. 467		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="LCLO58F01_r"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:		
	XhoI; The cDNA library was constructed from cells cultured		
	in a carbon stress acclimatized condition in which carbon		
	dioxide concentration in the bubbling gas was changed from		
	5% to 0.04%"		
BASE COUNT	108 a 149 c 142 g 68 t		
ORIGIN			
Query Match	5.9% Score 65.8; DB 10; Length 467;		
Best Local Similarity	56.3%; Pred. No. 5.1e-09;		
Matches 174; Conservative 0; Mismatches 117; Indels 18; Gaps 2;			
QY 50 TTGGCAATGTGCGCATGTTGCGCATGTGCTCCATCGCGTCACTGCAATGACCAATFACG 109			
DB 93 TCGGCAAGTGTGCGCGGTGGGACGACGACCGCATCGCCCTTCAAGCATGACACACAGG 152			
QY 110 GTAGCAGACAGTGCACACACGCTCAATCAATCAAGGCGCTGACACGCGTTGGCGCTG 169			
DB 153 ACACCCGGAATGTGAGGTGACCGGTGACACGAGTGAACAGTGCAGCGCGGCGCGG 212			
QY 170 AATATGCTCGTATCCCTTACCGAGATGAGGACGCGGAGAAAGCGTTCA----- 217			
DB 213 ACATGCTCCGATCATCATGTCAGGCAAGAGAGGCTGAGGCGTGCATGAAGATTCCGG 272			
QY 218 ---AATCATCAACACAGCAGGTTAACTGCGCTGCTGGTGGCTGCATCCACTTGCATATC 274			
DB 273 AGCAGCTGTTCAAGGACCAAGTACGACGCGCCCTGCTGGCGACATCCACTT---CCAGC 329			
QY 275 GCATTGGCTTGAAGTACGAGTAATGAGGCGTGCATGCTGCTCGATTAAACCTTGGCAATA 334			
DB 330 CCACGCGCGCATGATGTTGGCGGACGCTTTCAGAGACATCGTGTGAACCCCGCAACT 389			
QY 335 TCGGTAATG 343			
DB 390 TCGGTGACG 398			
RESULT 11			
AV641368			

LOCUS	AV641368	377 bp	mrna	linear	EST 15-DEC-2000
DEFINITION	AV641368 Chlamydomonas reinhardtii 5' CO2				Chlamydomonas reinhardtii
ACCESSION	AV641368				cdna clone HCl032d09_r 5', mRNA sequence.
KEYWORDS	AV641368.1	GI:10784696			
SOURCE	EST.				
ORGANISM	Chlamydomonas reinhardtii.				Chlamydomonas reinhardtii.
AUTHORS	Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,I., and Tabata,S.				Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,I., and Tabata,S.
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2				Generation of expressed sequence tags from low-CO2 and high-CO2
JOURNAL	adapted cells of Chlamydomonas reinhardtii				adapted cells of Chlamydomonas reinhardtii
MEDLINE	DNA Res. 7 (5), 305-307 (2000)				DNA Res. 7 (5), 305-307 (2000)
COMMENT	20539644				20539644
FEATURES	Contact: Erika Asamizu				Contact: Erika Asamizu
SOURCE	The First Laboratory for Plant Gene Research				The First Laboratory for Plant Gene Research
	Kazusa DNA Research Institute				Kazusa DNA Research Institute
	Yama 1532-3, Kisarazu, Chiba 292-0812, Japan				Yama 1532-3, Kisarazu, Chiba 292-0812, Japan
	Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.				Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
	Location/Qualifiers				Location/Qualifiers
	1..377				1..377
	/organism="Chlamydomonas reinhardtii"				/organism="Chlamydomonas reinhardtii"
	/strain="C9"				/strain="C9"
	/db_xref="taxon:3055"				/db_xref="taxon:3055"
	/clone="HCl032d09_r"				/clone="HCl032d09_r"
	/clone_lib="Chlamydomonas reinhardtii 5% CO2"				/clone_lib="Chlamydomonas reinhardtii 5% CO2"
	/note="vector: pluescriptII SK-, Site_1: EcoRI; Site_2: XhoI; The cdna library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"				/note="vector: pluescriptII SK-, Site_1: EcoRI; Site_2: XhoI; The cdna library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"
BASE COUNT	89 a	118 c	116 g	54 t	
ORIGIN					
Query Match	5.5%;	Score 62;	DB 10;	Length 377;	
Best Local Similarity	55.6%;	Pred. No. 6;	e-08;	Matches 150;	Conservative 0; Mismatches 105; Indels 15; Gaps 1;
Y	50	TTGGCAATGTCGCCGATGGTGGCGATGTCGTCCTCCATCGCCGTCGACAGTCATGCATATACG	109		
Db	102	TGGCAAGGTGCGGGTGGGAGCCAGCAGCCGATGCGCTTGAAGCAATGACACACACG	161		
Y	110	GTACACAGACGCTGGAACCAAGCTCAATCAATCAAGGCGCTGGAAAGCGCTTGCGCTG	169		
Db	162	AACCCGCAATGTGAGTGAACCGTGACCGTGACGACGAGTGAAGAAATGGCGGAGCCGCGGG	221		
Y	170	ATATGTCGCGTATCGCTACCGACGATGAGCGGCGCAGAGCGTTCA-----	217		
Db	222	ACATGTCGCGCATCACTGACGAGGCAAGAGAGGAGCTGAGCGCTGATGAATTCGCG	281		
Y	218	---AAGTCATCAACAGAGAGGTTAAAGCGCGCGTGGTGGCGTGCATCCACTTCGACTATC	274		
Db	282	AGCAAGCTTTGAAGAGCAAGTACGACGCGCCCTGGTGCGCGACATCCACTTCACGCCCA	341		
Y	275	GCATTGCGCTGAAGTAGACGAATACGGCG	304		
Db	342	CGGTGGCATATGTTGGCGGACGCGCTTCG	371		
RESULT 12					
LOCUS	AT1770308	380 bp	mrna	linear	EST 24-JAN-2000
DEFINITION	35 Mycobacterium anaerobic stationary phase library				Mycobacterium smegmatis cDNA, mRNA sequence.
ACCESSION	AT1770308				
KEYWORDS	AT1770308.1	GI:6742677			
SOURCE	EST.				
ORGANISM	Mycobacterium smegmatis.				Mycobacterium smegmatis.
	Mycobacterium smegmatis				Mycobacterium smegmatis
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
	Corynebacterineae; Mycobacteriaceae; Mycobacterium.				Corynebacterineae; Mycobacteriaceae; Mycobacterium.

REFERENCE 1 (bases 1 to 380)  
 Murgas-Oel, B., Tay, A. and Dick, T.  
 Upregulation of stress response genes and ABC transporters in  
 anaerobic stationary-phase *Mycobacterium smegmatis*  
 JOURNAL Mol. Gen. Genet. 262 (4-5), 677-682 (1999)  
 MEDLINE 20092472  
 COMMENT Contact: Murgas-Oel, B.  
 Mycobacterium Laboratory  
 Institute of Molecular and Cell Biology  
 30 Medical Drive, Singapore 117609, Republic of Singapore  
 Tel: 65 874 3011  
 Fax: 65 779 1117  
 Email: mcbomelmb.nus.edu.sg  
 Insert Length: 380 Std Error: 0.00  
 Seq primer: T3.

FEATURES  
 source Location/Qualifiers  
 1..380  
 /organism="Mycobacterium smegmatis"  
 /strain="mc2155"  
 /db\_xref="taxon:1772"  
 /clone\_lib="Mycobacterium anaerobic stationary phase library"  
 /lab\_host="E. coli XL1-Blue MRP"  
 /note="Vector: Lambda ZAP II; Bacilli were disrupted using a Mini bead beater (Biospec). RNA was isolated using the RNeasy protocol (Qiagen). Purified RNA was subjected to 2 rounds of digestion with RNase-free DNase I (Promega). DNase I was heat-inactivated at 75°C for 5 min. and removed by using RNeasy columns followed by phenol extraction and ethanol precipitation. The RNA preparations were confirmed to be free of genomic DNA contamination by carrying out PCR and RT-PCR using the histone-like protein gene hlp (Lee et al., 1998). cDNA was synthesized using random hexamer primers (Promega) and Stratagene's cDNA synthesis kit. cDNA fragments were ligated into Lambda ZAP II vector and packaged in vitro using Stratagene reagents."

BASE COUNT 68 a 130 c 121 g 61 t  
 ORIGIN

Query Match 5.0%; Score 56.4; DB 9; Length 380;  
 Best Local Similarity 53.7%; Pred. No. 3.7e-06;  
 Matches 117; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 47 ACCTTGGGATGTCGGATGCGATGCTGCTCCATCCCGACAGTCCATGACCAATTA 106  
 DB 163 ATGTGGGGCGGCGTGGCATCGGACGACGCCGATCGGGTGCATGTGACCA 222  
 QY 107 CGCGTACGACAGACGTCGAGCAACGTCATCAATCAAGCGCTGGAACGGCTGGCG 166  
 DB 223 CCAGAGCCACGACGTCACACTGACAGATGCCGCGCTGACCGCGCTGGCT 282  
 QY 167 GTGATATGCTCCGTATCCGTACGAGATGACGCGGACAGAGGCTTCAACTCATCA 226  
 DB 283 GTGACATCGTGGCTGGCGTCCCGTCAGAGAGACCGCGCTGGCGAGATCG 342  
 QY 227 AACGAGAGGTACGTCGCGCTGGCTGGCTGACATCCAC 264  
 DB 343 CGCGGACAGCCGATCGCGGTGATGCGGACATCCAC 380

RESULT 13  
 AV640429 303 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV640429 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
 DEFINITION CDNA clone HCL01b07\_1 5', mRNA sequence.  
 ACCESSION AV640429  
 VERSION AV640429.1 GI:10783757  
 KEYWORDS EST  
 SOURCE Chlamydomonas reinhardtii.  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE 1 (bases 1 to 303)  
 Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohnaya, K.,  
 Nakamura, Y. and Tabata, S.  
 Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of *Chlamydomonas reinhardtii*  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 MEDLINE 20539644  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yata 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
 Location/Qualifiers  
 1..303

FEATURES  
 source Location/Qualifiers  
 1..303  
 /organism="Chlamydomonas reinhardtii"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="Chlamydomonas reinhardtii 5% CO2"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"

BASE COUNT 77 a 92 c 92 g 42 t  
 ORIGIN

Query Match 4.8%; Score 54.2; DB 10; Length 303;  
 Best Local Similarity 55.6%; Pred. No. 1.5e-05;  
 Matches 104; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 50 TTGGGAATGTCGGATGCGATGTCGTCCTCCATCCCGCTACAGTCCATGACCAATACG 109  
 DB 99 TCGGCAAGTGGCGGCGGAGGACGACGACGATCCCTTCAGACATGACACACACG 158  
 QY 110 GTACGACAGACGTCGAGCAACGCTCATCAATCAAGCGCTGGAACCGCTGGCGCTG 169  
 DB 159 ACACCGCAATGTCAGTCAGTCAGTCCGTCGACAGTGAAGAATGCGCGGACGCGCGCG 218  
 QY 170 ATATCTCGCTGTATCCGTACGACGATGACGCGGACAGAGCGTCAAACTATCAAC 229  
 DB 219 ACATCTCGCATCATGACGTCAGGCAAGAGAGGCTGAGCGCTCATGACATTCGCG 278  
 QY 230 AGCAGCT 236  
 DB 279 AGCAGCT 285

RESULT 14  
 AW282425 597 bp mRNA linear EST 09-FEB-2000  
 LOCUS AW282425 829011H05.x1 829 - silk infected with Fusarium Zea mays cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION AW282425  
 VERSION AW282425.1 GI:6672259  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 597).

REFERENCE 1 (bases 1 to 597).  
 Walbot, V.  
 Maize ESTs from various cDNA libraries sequenced at Stanford  
 University  
 Unpublished (1999)  
 CONTACT Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu

Plate: 829011 Row: H Column: 05.

## FEATURES

## Source

Location/Qualifiers  
1. 597

/organism="zea mays"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone\_id="829 - Silk infected with Fusarium"

/tissue\_type="silk"

/dev\_stage="adult"

/lab\_host="DH10p"

/note="Organ: silk; Vector: pBluescript II XR; Site: 1; XhoI; Site: 2; EcoRI; cDNA library of silks infected with 1 microfilter of 500,000 spores/ml solution of Fusarium graminearum DAOM 180378. Prepared by Sharon Allard of Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada using Stratagene cDNA synthesis kit. Silk was harvested at 72 hours p.i."

BASE COUNT 157 a 139 c 187 g 113 t 1 others  
ORIGIN

## Query Match

Best Local Similarity 54.0%; Pred. No. 3,1e-05; Length 597;  
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 26 GTGAAATATCAACGATTTACCTGGGAATGTCGGATGGCGATGTCCTCCATCG 85  
DB 393 GGAGAGAAACATCGAATCTGATGTGGGAATGTCACCTGCGAGATCATCCATTA 452  
QY 86 CCGTACAGTCATGACCAATACGCGTACGACAGCTGGAAGCAAGGTCATCAATCA 145  
DB 453 GGATTCACACCATGACACCTTCAGATACCAAGATGTTGCAAAAACATGAGAGGTGA 512  
QY 146 AGCGCTGGAAGCGGTTGGCGCTGATATCGCTGTATCCGTACCGACGATGACCGG 205  
DB 513 TGAGGATAGCAGATTAAGAGAGCTGATCTGTAGTAATACAGTCCAGGATGGAAGAG 572  
QY 206 CAGAGCGTTCAACATCATCA 227  
DB 573 CTGATGCTGCTTTGAGATCAA 594

## RESULT 15

## BG525868

LOCUS BG525868 728 bp mRNA linear EST 16-NOV-2001

DEFINITION 53-47-1-2 r Stevia field grown leaf cDNA Stevia rebaudiana cDNA 5',  
mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## EST

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CONTACT

## GENOMICS

## AGRICULTURE

## TELEPHONE

## FAX

## EMAIL

## SEG PRIMER

## LOCATION

## QUALIFIERS

## ORGANISM

## CULTIVAR

## LANDRACE

## SOURCE

BASE COUNT 234 a 120 c 167 g 204 t 3 others  
ORIGIN

## Query Match

Best Local Similarity 52.7%; Pred. No. 4e-05; Length 728;  
Matches 175; Conservative 0; Mismatches 139; Indels 18; Gaps 2;

QY 26 GTGAAATATCAACGATTTACCTGGGAATGTCGGATGGCGATGTCCTCCATCG 85  
DB 163 GGAGAGAAACATCGAATCTGATGTGGGAATGTCACCTGCGAGATCATCCATTA 222  
QY 86 CCGTACAGTCATGACCAATACGCGTACGACAGCTGGAAGCAAGGTCATCAATCA 145  
DB 223 GAATTCACACCATGACACCTTCAGATACCAAGATGTTGCAAAAACATGAGAGGTGA 282  
QY 146 AGCGCTGGAAGCGGTTGGCGCTGATATCGCTGTATCCGTACCGACGATGACCGG 205  
DB 283 TGCAATAGCTGACAGAGAGAGCTGATCTGTAGTAATACAGTCCAGGATGGAAGAG 342  
QY 206 CAGAGCGTT-----CAACTCATCAACAGCAGGTTACGTGCCCTGG 250  
DB 343 CAGATGATGTTTGAATTAACACACCTGTTGCAAAAACATTAATACATCTCTGG 402  
QY 251 TGCGTACATCCACTGCACTATGCGATGCGCTGGAAGTAGCGGAATACCGCTGAT 310  
DB 403 TGGCAGACATTCATTTGACACCTTCAGTGCATGCGCGGTGCTGAT---GTTTGATA 459  
QY 311 GTCTGCTATTAACCTCGGCAATATCGTAT 342  
DB 460 AATTCGTGTCACCCCTGGAATTTTCTGAT 491

Search completed: May 10, 2003, 16:41:00  
Job time: 1906 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 15:09:17 ; Search time 78 Seconds  
(without alignments)  
4399.632 Million cell updates/sec

Title: US-09-921-992-3  
Perfect score: 1119  
Sequence: 1 atgcataaccaggtcccaat.....ttcagcaggttgaaataa 1119

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338361 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCBUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1010	90.3	1010	2	US-08-827-190-8
2	1010	90.3	1010	4	US-09-170-187-8
3	631.6	56.4	1345	2	US-08-827-190-4
4	631.6	56.4	1345	4	US-09-170-187-4
5	619.8	55.4	1079	2	US-08-827-190-7
6	619.8	55.4	1079	4	US-09-170-187-7
7	534.6	47.8	886	2	US-08-827-190-10
8	534.6	47.8	886	4	US-09-170-187-10
9	269	24.0	886	2	US-08-827-190-9
10	269	24.0	886	4	US-09-170-187-9
11	55.4	5.0	1083	4	US-09-221-017B-309
12	40.6	3.6	1206	4	US-09-153-599A-2
13	40.6	3.6	2299	4	US-09-153-599A-1
14	38.4	3.4	1665	4	US-09-221-017B-1017
15	38.2	3.4	4411529	4	US-09-103-840A-1
16	35.4	3.2	583	4	US-08-998-416-1089
17	35.4	3.2	737	4	US-08-998-416-1039
18	35	3.1	1383	3	US-09-400-742-7
19	35	3.1	1383	3	US-08-618-651A-7
20	35	3.1	1383	4	US-09-215-252-6
21	34	3.0	8910	3	US-08-779-764A-1
22	34	3.0	4411529	4	US-09-103-840A-1
23	33.8	3.0	44377	2	US-08-804-227C-7
24	33.8	3.0	44377	2	US-08-804-198-1
25	33.6	3.0	606	2	US-08-432-871C-48
26	33.6	3.0	606	2	US-09-270-956-48
27	33.4	3.0	1660	4	US-09-297-468-1

C 28	33.4	3.0	23673	4	US-09-773-816-1	Sequence 1, Appl
C 29	32.8	2.9	71969	4	US-09-443-501A-2	Sequence 2, Appl
C 30	32.4	2.9	5135	3	US-08-369-822C-9	Sequence 9, Appl
C 31	32.4	2.9	5135	3	US-08-582-776C-9	Sequence 9, Appl
C 32	32.4	2.9	5135	3	US-08-434-818-9	Sequence 9, Appl
C 33	32.4	2.9	5698	1	US-08-761-258-11	Sequence 11, Appl
C 34	32.4	2.9	5698	2	US-08-977-306-11	Sequence 11, Appl
C 35	32.4	2.9	8910	3	US-08-369-822C-19	Sequence 19, Appl
C 36	32.4	2.9	8910	3	US-08-582-776C-19	Sequence 19, Appl
C 37	32.4	2.9	8910	3	US-08-434-818-19	Sequence 19, Appl
C 38	32.4	2.9	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 39	31.8	2.8	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 40	31.6	2.8	43280	2	US-08-804-227C-1	Sequence 1, Appl
C 41	31.4	2.8	1912	1	US-08-270-013B-1	Sequence 1, Appl
C 42	31.4	2.8	1912	1	US-08-838-418-1	Sequence 1, Appl
C 43	31.2	2.8	1778	5	PCT-US91-02958-9	Sequence 9, Appl
C 44	31.2	2.8	3934	3	US-09-226-568-18	Sequence 18, Appl
C 45	31.2	2.8	3946	1	US-08-077-848A-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-827-190-8  
Sequence 8, Application US/08827190  
Patent No. 5858367  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1010 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-827-190-8  
Query Match 90.3% Score 1010; DB 2; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 5.8e-309; Indels 0; Gaps 0;  
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 ATGCATACACAGGCTCCATTCAGTAGAAGTAACACAGCTATTACGTTGAGATGTG 60  
Db 1 ATGCATACACAGGCTCCATTCAGTAGAAGTAACACAGCTATTACGTTGAGATGTG 60  
OY 61 CCGATTGGCGATGTGTCTCCATCGCCGTACAGTCCATGACCAATATACGATACGACAGAC 120

|||||  
Db 61 CCGATTGGCGATGGTGTCTCCATCGCCGTACAGTCCATGACCAATATACGCGTACAGACAGAC 120  
QY 121 GTGGAAGCAACGGGTCAATCAATCAAGCGCGTGGAAAGCGCTGGGCGTGTATGTCCT 180  
Db 121 GTGGAAGCAACGGGTCAATCAATCAAGCGCGTGGAAAGCGCTGGGCGTGTATGTCCT 180  
QY 181 GTATCCGTACCGAGATGAGAGCGGCGAGAACGCTTCAACTCATCAACAGCAGGTTAAC 240  
Db 181 GTATCCGTACCGAGATGAGAGCGGCGAGAACGCTTCAACTCATCAACAGCAGGTTAAC 240  
QY 241 GTGCGCGTGTGTGCTGACATCACTTCGACTATTCGACTTCGCGTGAAGTACCGGATAC 300  
Db 241 GTGCGCGTGTGTGCTGACATCACTTCGACTATTCGACTTCGCGTGAAGTACCGGATAC 300  
QY 301 GGGGTGATGTGTGCTGATTAACCCGTGGCAATATGATGATGAAGCGCTATTCGATG 360  
Db 301 GGGGTGATGTGTGCTGATTAACCCGTGGCAATATGATGATGAAGCGCTATTCGATG 360  
QY 361 GTGCTGACTGTGCGCGCGATTAACCAATTCGATCCGATTTGGCGTTAACGCGGATCG 420  
Db 361 GTGCTGACTGTGCGCGCGATTAACCAATTCGATCCGATTTGGCGTTAACGCGGATCG 420  
QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACCGACGCGGCGGCTTGTGGAATCT 480  
Db 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACCGACGCGGCGGCTTGTGGAATCT 480  
QY 481 GCGATGGGTATGTGATCTATCTGATCGATCGGCTGATGATGATGATGATGATGATG 540  
Db 481 GCGATGGGTATGTGATCTATCTGATCGATCGGCTGATGATGATGATGATGATGATG 540  
QY 541 AAAGCGTCTGACGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
Db 541 AAAGCGTCTGACGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
QY 601 CACCGCTGATCTGCGGATGATCCGGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Db 601 CACCGCTGATCTGCGGATGATCCGGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
QY 661 GCGATGGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 661 GCGATGGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
QY 721 GCGCGCGATCCGCTGCAAGAGATCAAAAGTGGTTGCAATTTTGAATCGCTGATC 780  
Db 721 GCGCGCGATCCGCTGCAAGAGATCAAAAGTGGTTGCAATTTTGAATCGCTGATC 780  
QY 781 CGTTCGAGAGGATCACTCATCGCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 781 CGTTCGAGAGGATCACTCATCGCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
QY 841 ATCGTACCGTAAACGCGCTGAGAGCAAGCGCTGGAAGATATCATCTCGATGAGAGT 900  
Db 841 ATCGTACCGTAAACGCGCTGAGAGCAAGCGCTGGAAGATATCATCTCGATGAGAGT 900  
QY 901 TCGATTATGCGCTGCGTGTGGAATGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 901 TCGATTATGCGCTGCGTGTGGAATGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
QY 961 ACCGAGGCGCAAGAAAGCGGCTCTATGAAGATGCGGTGGCAAGA 1010  
Db 961 ACCGAGGCGCAAGAAAGCGGCTCTATGAAGATGCGGTGGCAAGA 1010

## RESULT 2

US-09-170-187-8  
Sequence 8, Application US/09170187  
Patent No. 6383745

GENERAL INFORMATION:  
APPLICANT: Ralcher, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,187  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEO ID NO: 8  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1010 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-09-170-187-8

Query Match 90.3%; Score 1010; DB 4; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 5,8e-309;  
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATATACCAAGGCTCAATTCACGTAAGAAATCAACAGCTATTCGTTGGGAATGTG 60  
Db 1 ATGATATACCAAGGCTCAATTCACGTAAGAAATCAACAGCTATTCGTTGGGAATGTG 60  
QY 61 CCGATTGGCGATGGTGTCTCCATCGCCGTACAGTCCATGACCAATATACGCGTACAGAC 120  
Db 61 CCGATTGGCGATGGTGTCTCCATCGCCGTACAGTCCATGACCAATATACGCGTACAGAC 120  
QY 121 GTGGAAGCAACGGGTCAATCAATCAAGCGCGTGGAAAGCGCTGGGCGTGTATGTCCT 180  
Db 121 GTGGAAGCAACGGGTCAATCAATCAAGCGCGTGGAAAGCGCTGGGCGTGTATGTCCT 180  
QY 181 GTATCCGTACCGAGATGAGAGCGGCGAGAACGCTTCAACTCATCAACAGCAGGTTAAC 240  
Db 181 GTATCCGTACCGAGATGAGAGCGGCGAGAACGCTTCAACTCATCAACAGCAGGTTAAC 240  
QY 241 GTGCGCGTGTGTGCTGACATCACTTCGACTATTCGACTTCGCGTGAAGTACCGGATAC 300  
Db 241 GTGCGCGTGTGTGCTGACATCACTTCGACTATTCGACTTCGCGTGAAGTACCGGATAC 300  
QY 301 GGGGTGATGTGTGCTGATTAACCCGTGGCAATATGATGATGAAGCGCTATTCGATG 360  
Db 301 GGGGTGATGTGTGCTGATTAACCCGTGGCAATATGATGATGAAGCGCTATTCGATG 360  
QY 361 GTGCTGACTGTGCGCGCGATTAACCAATTCGATCCGATTTGGCGTTAACGCGGATCG 420  
Db 361 GTGCTGACTGTGCGCGCGATTAACCAATTCGATCCGATTTGGCGTTAACGCGGATCG 420  
QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACCGACGCGGCGGCTTGTGGAATCT 480  
Db 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACCGACGCGGCGGCTTGTGGAATCT 480  
QY 481 GCGATGGGTATGTGATCTATCTGATCGATCGGCTGATGATGATGATGATGATGATG 540  
Db 481 GCGATGGGTATGTGATCTATCTGATCGATCGGCTGATGATGATGATGATGATGATG 540



```

Db 481 GCCATGCGTATGTTGATCATCTCATGCGCTGAACTTCGATGAGTCAAGTCAGCGTG 540
Oy 541 AAAGCGTCTGACGCTCTTCCTGCTGTTGAGTCTTAATCGTTTCTGGCAAAACAGATCGAT 600
Db 541 AAAGCGTCTGACGCTCTTCCTGCTGTTGAGTCTTAATCGTTTCTGGCAAAACAGATCGAT 600
Oy 601 CAGCGCTTGACATCTGGGGATCACGCAAGCGGTGTGCGGACGCGGGGAGTAAATCC 660
Db 601 CAGCGCTTGACATCTGGGGATCACGCAAGCGGTGTGCGGACGCGGGGAGTAAATCC 660
Oy 661 GCCATTGTTTGGTCTGCTGCTGTCTGAAGCATCGGCAACAGCTGCGCTATCGCTG 720
Db 661 GCCATTGTTTGGTCTGCTGCTGTCTGAAGCATCGGCAACAGCTGCGCTATCGCTG 720
Oy 721 GCGGCGCATCCGCTGGAAGAGATCAAAAGTGGTTTCGATATTTTGAATCGTGGCTATC 780
Db 721 GCGGCGCATCCGCTGGAAGAGATCAAAAGTGGTTTCGATATTTTGAATCGTGGCTATC 780
Oy 781 CGTTCGCGAGGATCAACTTCATCGCTGCGCGACCTGTTGCGCTCAGGAATTTGATGT 840
Db 781 CGTTCGCGAGGATCAACTTCATCGCTGCGCGACCTGTTGCGCTCAGGAATTTGATGT 840
Oy 841 ATCGGTACCGTTAAGCGCTGAGACCAAGCCCTGGAAGATATCATCTCCGATGGACGT 900
Db 841 ATCGGTACCGTTAAGCGCTGAGACCAAGCCCTGGAAGATATCATCTCCGATGGACGT 900
Oy 901 TCGATTATGCGCTGCTGCTGATGAGGCCAGGTGAGAGCGCTGTTCTACACTGCGCGTC 960
Db 901 TCGATTATGCGCTGCTGCTGATGAGGCCAGGTGAGAGCGCTGTTCTACACTGCGCGTC 960
Oy 961 ACCGCGCGCACAGAAAAGCGCCTCTATGAAGATGCGGTGCGCAAGA 1010
Db 961 ACCGCGCGCACAGAAAAGCGCCTCTATGAAGATGCGGTGCGCAAGA 1010

```

RESULT 3  
US-08-827-190-4  
Sequence 4, Application US/08827190  
Patent No. 5858367

## GENERAL INFORMATION:

APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:

## CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1345 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 175..1272  
OTHER INFORMATION: /gene="aarc"  
OTHER INFORMATION: /note="Similar to E. coli Gcpe protein listed by GenBank."  
US-08-827-190-4

Query Match  
Best Local Similarity 56.4%; Score 631.6; DB 2; Length 1345;  
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;

```

Oy 1 ATGATAACAGAGCTCCATTCACGTAGAAATCAACACGATTTTACGTGGGAATG 60
Db 175 ATGATAATGAATCAACCGATTAAGAGCTAAATCCACCGAATTTATAGTAACGTG 234
Oy 61 CCGATTGGGATGATGCTCCATCCGCGTACAGTCCATACCAATACGGGTACGACAGC 120
Db 235 CCTATTGGGATGATGCTCCATCCGCGTACAGTCCATACCAATACGGGACGAGGAT 294
Oy 121 GTGGAAGCAACGCTCAATCAATCAAGGCGCTGGAACGCGTGGCCTGATATCGTCGT 180
Db 295 GTTGAAGCACCTGCTGGGCAATCCATCATCTTGAAGCTGTAGTGTATATCGTCCG 354
Oy 181 GTATCCGTACGACGATGAGACCGCGCAGAGAGCGTTCAACATCAACAGCAGTTAATC 240
Db 355 GTGTCTGTTCATACGATGATGATGACGAGAGAGCGTTAATTAATTAAGCAGCGCTGAT 414
Oy 241 GTGCGCTGTGTGCTGACATTCACCTTGCATCTGATTCGATTGCGGTGAAGTACGAAATC 300
Db 415 GTGCAATTTGTGCGGATTTTCACCTTGTACCTGATTCGATGCGATGAAGTGGCTGAATAT 474
Oy 301 GCGGTGATTTGTGCTGCTGATTAACCTGGCAATATGATGATGAAGAGGATATTCGATG 360
Db 475 GGTGTGACTGCTTACGAAATTAACCGAGTAATATGCGAGTGAAGAGCTATTCGCCAA 534
Oy 361 GTGTGACTGTGCGCGCATTAACCATTCGATCCGATTATGCGCTTAAACGCGGATG 420
Db 535 GTGTGATGATGCTGCTGATCACAACATTCGATTCGATATGAGGCTCAATGCGGCTCA 594
Oy 421 CTGGAAGAAATCTGCAAGAAAAGTATGCGCAACGACGCGGAGCGCTTCTGCAATCT 480
Db 595 CTGGAAGAAATCTGCAAGAAAAGTATGCGCAACGACGCGGAGCGCTTCTGCAATCT 654
Oy 481 GCCATGCTGATGATCATCTCATGCGCGCAATTCGATGATCAAGTCACACGTG 540
Db 655 GCATGCGACATGTTGATATCTTGGACAGCTGAATTTGATCAGTCAAGTCACTGT 714
Oy 541 AAAGCTCTGAGCTTCTCCTGCTGTTGAGTCTTATCGTTTGTGCAAAACAGATCGAT 600
Db 715 AAAGCTGAGATGTTCTTCTGCGCTGCTTATCGTTTATGGCGCAAAATTTGAT 774
Oy 601 CAGCGCTGACATCTGGGGATCACCGAAGCGGTGTGCGGACGCGGCGAGTAAATCC 660
Db 775 CAACCACTTCACCTCGTATTCAGAAAGCGGGGCTGCTTCTGTTCAAGTAATCA 834
Oy 661 GCCATTGTTTGGTCTGCTGCTGTGTAAGGATGCGGAGCAACGCTGGGCGGATGCTG 720
Db 835 GCATTTGCTTGTGATGTTGTTGGCTGAAGGATGCGGATACGTTACGATCTCTACTC 894
Oy 721 GCGGCGCATCCGCTGGAAGAGATCAAAAGTGGTTTCGATATTTTGAATCGTGGCTATC 780
Db 895 GCGGCGCATCCGCTGGAAGAGATCAAAAGTGGTTTCGATATTTTGAATCGTGGCTATC 954
Oy 781 CGTTCGCGAGGATCAACTTCATCGCTGCGGACCTGTTGCGCTGAGGAATTTGATGT 840
Db 955 CGCTACGCGGATCAACTTTATGCTTCCCAACCTGTTCAAGCCCAAAATTTGATGTG 1014
Oy 841 ATCGGTACCGTTAAGCGCTGAGCAACGCTGGAAGATATCATCTCCGATGGACGT 900
Db 1015 ATTTGATCGGTAATATGCTTTGGAGACGCGCTCTGGAAGATATTTATCAGCGCGATGATGTC 1074
Oy 901 TCGATTATGCGCTGCTGCTGATGAGGCCAGGTGAGGCGCTGTTCTACACTGCGCGTC 960

```

Db 1075 TCTATTATTGTTGTGTAAGGAATGGCCGCGGTGAAGCCGAGGTTCTACTTTAGGCTG 1134  
QY 961 ACCGGCGGCAACAAGAGGGGCTCTATGAAGATGGCGTGC---AAAGACCGCTG 1017  
Db 1135 GCTGGCGCGAAACCAAGGTTTCTATGAAGATGGCTTGCAGAAAAGAGCGTTT 1194  
QY 1018 GACAAACAGATATGATGCGAGCTGGAGCAACGATTCGTGGCAAGCCAGTCAGCTG 1077  
Db 1195 GATAATGACATATTTATTTGATCAGCTTGAGGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254  
QY 1078 GACGAA 1083  
Db 1255 GATGAA 1260

RESULT 4  
US-09-170-187-4  
Sequence 4, Application US/09170187  
Patent No. 6383745  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/170,187  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-0410  
TELEFAX: (415) 397-8398

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1345 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 175..1272  
OTHER INFORMATION: /gene="aarc"  
OTHER INFORMATION: /note="Similar to E. coli Gcpe protein listed by GenBank."

US-09-170-187-4

Query Match 56.03; Score 631.6; DB 4; Length 1345;  
Best Local Similarity 74.33; Pred. No. 2e-189;  
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;

QY 1 ATGCATACAGGCTCCATTCATGAGAAATCAACAGATTTACGTTGGAAATG 60  
Db 175 ATGCATATATCATCAGGAAAGAAAGAGAGATTAATCCAGCCGAAATTTATGATGAGTAC 234

QY 61 CCGATTGGCGATGCTGCTCCATTCGCGGTACAGTCAACCAATACCGTACGACAG 120  
Db 235 CCGATTGGCGATGCTGCTCCATTCGCGGTACAGTCAACCAATACCGTACGACAG 294  
QY 121 GTGCAAGCAACGCGATCAATCAATCAAGCGCTGGAGCGCTGGCGCTGATGCTG 180  
Db 295 GTTGAAGCAACGCGATCAATCAATCAAGCGCTGGAGCGCTGGCGCTGATGCTG 354  
QY 181 GTATCCGTACGAGATGAGCGCGGAGAGCGCTTCAACTCAATCAACCAAGCTTAC 240  
Db 355 GTGCTCTTCTTACGATGATGAGCGAGAGCGCTTCAATCAATCAACCAAGCTTAC 414  
QY 241 GTGCGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 300  
Db 415 GTGCGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 474  
QY 301 GCGCTGATGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 360  
Db 475 GGTGTTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 534  
QY 361 GTGCTGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 420  
Db 535 GTGCTGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 594  
QY 421 CTGCAAAAGATCTGCAAGAAAGTATGCGAACCAGCGCGAGCGCTTCTGGAATCT 480  
Db 595 CTGCAAAAGATCTGCAAGAAAGTATGCGAACCAGCGCGAGCGCTTCTGGAATCT 654  
QY 481 GCGATGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 540  
Db 655 GCGATGCTGCTGCTGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTAC 714  
QY 541 AAAGCGTCTGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
Db 715 AAAGCGTCTGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774  
QY 601 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Db 775 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 834  
QY 661 GCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 835 GCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894  
QY 721 GCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Db 895 GCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954  
QY 781 GCTTGGGAGGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTCTGGA 840  
Db 955 GCTTGGGAGGATCAATCAATCAAGCGCTTCAACTCAATCAACCAAGCTTCTGGA 1014  
QY 841 ATGCGTACGCTTAAAGCGCTGAGCAACGCGCTGGAAGATATCACTCGATGAGCTT 900  
Db 1015 ATGCGTACGCTTAAAGCGCTGAGCAACGCGCTGGAAGATATCACTCGATGAGCTT 1074  
QY 901 TCGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 1075 TCGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1134  
QY 961 ACCGGCGGCAACAAGAGGGGCTCTATGAAGATGGCGTGC---AAAGACCGCTG 1017  
Db 1135 GCTGGCGCGAAACCAAGGTTTCTATGAAGATGGCTTGCAGAAAAGAGCGTTT 1194  
QY 1018 GACAAACAGATATGATGCGAGCTGGAGCAACGATTCGTGGCAAGCCAGTCAGCTG 1077  
Db 1195 GATAATGACATATTTATTTGATCAGCTTGAGGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254  
QY 1078 GACGAA 1083  
Db 1255 GATGAA 1260



ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE//DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-09-170-187-7

Query Match 55.4%; Score 619.8; DB 4; Length 1079;  
Best Local Similarity 74.9%; Pred. No. 9.4e-186;  
Matches 790; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

QY 17 CAATTCAGCTAGAAAATCAACACGATTTACGTTGGAAATGCGCGATTTGGCGATG 76  
DB 4 CTATTAGCGCTCGATGACAAAATTTATGTGGAAATGTACCAATTTGGTGGGATG 63  
QY 77 CTCCATCGCGCTACATTCATGACCAATACGCTACGACAGAGCTGCAACAGCTCA 136  
DB 64 CGCTATTGCGCTGCAATCAATGCAATATCTCGACCTGATGAGAGCGACAGTTG 123  
QY 137 ATCAATCAAGGCGCTGGAACGCGTGTGCGCTGATGCTCGCTATCCGTAACCGACA 196  
DB 124 CTCAATTAATCATTTGAAGAGTGTGTGAGATATTTGCTGCTATCTGTTCCAAACA 183  
QY 197 TGGACGGGCAAGAGCTTAAACTATCAACAGAGTTAACGCTCCGCTGCTGCTG 256  
DB 184 TGGATGCTGCGGAAGCTTTAAACAAATTAACAAAGTAAAGTATCCGCTGACAG 243  
QY 257 ACATCCATTCGACTACGATGCGCTGGAAGTAGCGGAATACGCGCTGATGCTGCTG 316  
DB 244 ATATCATTTGCACTACGATGCGCTTAAAGTCGCAATATGAGTGGATTTGTTAC 303  
QY 317 GATTAACCTGCGCAATATCGTAATGAGAGGCTATTCGATGCTGCTGCTGCTGCTG 376  
DB 304 GTATCATCTCGGAACATTTGCTGTAATGATGCGCTGCGCTGCTGCTGCTGCTGCTG 363  
QY 377 GCGATTAACATTCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436  
DB 364 GAGACAAAATATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423  
QY 437 AAGAAAGTATGGGCAACGCGCGAGCGCTGCTGGAATGCTGCTGCTGCTGCTGCTGCTG 496  
DB 424 AAGAAAATATGCGAAGCAACGCGCAAGCTTGTAGATCCGCTGCTGCTGCTGCTGCTG 483  
QY 497 ATCATTCGATCGCTGCAATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556  
DB 484 AATTTCTAGATCTCTTAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543  
QY 557 TCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 616  
DB 544 TCTTACGCGGTGATCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603  
QY 617 GATTCACCGAAGCGCGTGGCGGCGAGCGGCGGCGAGTAAATCCGCTGCTGCTGCTGCTG 676  
DB 604 GATTTACAGAGAGGCTGGCGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663  
QY 677 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736  
DB 664 TGTATTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723  
QY 737 AAGAGTCAAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 796  
DB 724 AAGAAATCAAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783  
QY 797 ACTTCATCGCTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856

DB 784 ACTTATTTGCTTGGCCCAACGCTTCTCGCCAGAAATTTGATGATGATGATGATGATGATG 843  
QY 857 CGGTGAGCAACGCGCTGGAAGATATCATCTCGCATGAGAGTTCGATATGCGTGGC 916  
DB 844 CGGTAGAACAGCGCTTGAAGATATTTATACCAATGATGATGATGATGATGATGATGATG 903  
QY 917 TGTGAATGCGCCAGGTGAGGCGCTGCTTCTACACTGCGCGCTGACCGCGGCAACAGA 976  
DB 904 TAGTGAATGCTGCTGCGGAGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963  
QY 977 AAGCGGCTCTATGAAGATGCGCTGCGC---AAGACGCTTGGACAAACAGATATGA 1033  
DB 964 AAGCGGCTTATTTATCTTACGAGGAGAGCCAAAAGAGCGCTTTGATTAACAGATATAG 1023  
QY 1034 TCGACGAGCTGGAAGCAGCATTGCTGCGAAGCC 1068  
DB 1024 TGAACCAATTAAGAGCAAAATTTGCTGCAAGCTC 1058

RESULT 7  
US-08-827-190-10/C  
Sequence 10, Application us/08827190  
Patent No. 5858367  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE//DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 886 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-827-190-10

Query Match 47.8%; Score 534.6; DB 2; Length 886;  
Best Local Similarity 75.3%; Pred. No. 6.7e-159;  
Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 83 TCCGCTACAGTCCATGACCAATACGCTGACGACAGACGCTGAGCAACGCTCAATCAAA 142  
DB 886 TTTCTTTCCATCTATGAGAAATATTTTCTGAGATGTTGAGATGTTGAGATGTTGAGATG 827  
QY 143 TCAAGGCGCTGGAACGCGTGGCGCTGATATCGCTGATATCGTACGACGATGAGAG 202  
DB 826 TCCATACACTTGAAGCGGTGATGTTGATATGCTCCGCTGCTGCTGCTGCTGCTGCTGCTG 767

QY 203 CGGAGAGAGCTTCAACTCATCAACAGAGGTTAACGTCGCGCTGGTGCATCC 262  
 Db 766 CAGCAGAGAGCTTTAAATTAATTAAGCAGCGCGTGAATGATGCGATATTC 707  
 QY 263 ACTTGACTATCGCATTCGCGTGAAGAGGTAATGAGGCTGATGTCGCTATTA 322  
 Db 706 ACTTGTACATCCGTAATCGCATGAAGTGCGTAATGAGTGTACTGCTACCAATTA 647  
 QY 323 ACCGTGCAATTCGTAATGAGAGAGCTATTCGATGTCGTCGCGCGCAATA 382  
 Db 646 ACCGAGTAATATTCGCGCATGAGAGCGTATTCGCGCACTGCTATAGTGTCTATC 587  
 QY 383 AAAACATTCCGATCCGATTTGGCGTTAAACGCCGATCCGTGAAAAAGATCGAAGAA 442  
 Db 586 ACAACATTCCTATCCGTAATGAGGCTCAATGCGGCTCACTGAAAAAGATATCCAAGAA 527  
 QY 443 AGTATGGGAACCGAGCGCGCGAGGCTGCTGATATCCGATGCTATGATCATC 502  
 Db 526 AATACGCTGAGCAACACCTGAAAGCTGTTGATCAAGCAATGCGACATGTTGATATCT 467  
 QY 503 TCGATCGCTGAACTTCATCATCAAGTCAAGCGCTGAAAGCGCTGACGCTTCCTCG 562  
 Db 466 TGGACAGGCTGAATTCGATCATGATCAAGTCAAGTCAAGCGTCAAGTCTTCTTG 407  
 QY 563 CTGTTGACTCTTATTCGTTGCTGCGAATAACAGATCATGACGCTGATCGGAGATCA 622  
 Db 406 CCGTGGCTCTTATTCGTTGCTGCGAATAACAGATCATGACGCTGATCGGATTA 347  
 QY 623 CCGAAGCGGCTGCTGCGCGAGCGAGTAAATCCGCGCTGCTTACGTCGCTGCTG 682  
 Db 346 CAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287  
 QY 683 TGTCTGAAGCATTCGCGAGCAGCGTATCGCTGCGCGCGCTGCGCGAGAGA 742  
 Db 286 TGGCTGAAGCATTCGCGAGCAGCGTATCGCTGCGCGCGCTGCGCGAGAGA 227  
 QY 743 TCAAGTGGCTTTCATATTTGAATGCGTCCGATCCGTTGCGGAGGATCAACTTCA 802  
 Db 226 TGAAGTGGCTTTCATATTTGAATGCGTCCGATCCGTTGCGGAGGATCAACTTCA 167  
 QY 803 TGGCTGGCGCGAGCTGCTGCGAGATTTGATGTTATGCTGCGTAAAGCGGCTG 862  
 Db 166 TGGCTGGCGCGAGCTGCTGCGAGATTTGATGTTATGCTGCGTAAAGCGGCTG 107  
 QY 863 ACCAAGCGCTGGAATATCATCATCCGATGAGCGTTCGATTTACGCTGCTGCTG 922  
 Db 106 ACCAAGCGCTGGAATATCATCATCCGATGAGCGTTCGATTTACGCTGCTGCTG 47  
 QY 923 ATGCGCCAGGTAGGCGCTGTTTCTACACTCGCGCTACCGCGC 967  
 Db 46 ATGCGCCAGGTAGGCGCTGTTTCTACACTCGCGCTACCGCGC 2

RESULT 8  
 ; Sequence 10, Application US/09170187  
 ; Patent No. 6383745  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ralther, Phillip N.  
 ; TITLE OF INVENTION: Methods for Screening For Antimicrobials  
 ; TITLE OF INVENTION: Utilizing aatc And Compositions Thereof  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/170,187  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/827,190  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Carroll, Peter G.  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: CASE-02443  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 886 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-170-187-10  
 ;  
 ; Query Match 47.8%; Score 534.6; DB 4; Length 886;  
 ; Best Local Similarity 75.3%; Pred. No. 6.7e-159;  
 ; Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;  
 ;  
 QY 83 TCGCGCTGACGTCATTCAGCAATGAGCGGTAGCAGAGCTGCAAGCAACGGTCATCA 142  
 Db 886 TTGCTGTCCTCAATCTATGAGCAATGAGCGGTAGCAGAGCTGCAAGCAACGGTCATCA 827  
 QY 143 TCAAGCGCTGGAAGCGCTGCGCTGATGTCGCTGATCCGTACCGAGATGAGC 202  
 Db 826 TCCATATCACTTGAAGGTAGGTGTTGATATGTCGCGCTGCTGCTGATGAGATG 767  
 QY 203 CGGAGAGAGCTTCAACTCATCAACAGAGGTTAACGTCGCGCTGGTGCATCC 262  
 Db 766 CAGCAGAGAGCTTTAAATTAATTAAGCAGCGCGTGAATGAGTGTACTGCTACCAATTA 707  
 QY 263 ACTTGACTATCGCATTCGCGTGAAGAGGTAATGAGGCTGATGTCGCTATTA 322  
 Db 706 ACTTGTACATCCGTAATCGCATGAAGTGCGTAATGAGTGTACTGCTACCAATTA 647  
 QY 323 ACCGTGCAATTCGTAATGAGAGAGCTATTCGATGTCGTCGCGCGCAATA 382  
 Db 646 ACCGAGTAATATTCGCGCATGAGAGCGTATTCGCGCACTGCTATAGTGTCTATC 587  
 QY 383 AAAACATTCCGATCCGATTTGGCGTTAAACGCCGATCCGTGAAAAAGATCGAAGAA 442  
 Db 586 ACAACATTCCTATCCGTAATGAGGCTCAATGCGGCTCACTGAAAAAGATATCCAAGAA 527  
 QY 443 AGTATGGGAACCGAGCGCGCGAGGCTGCTGATATCCGATGCTATGATCATC 502  
 Db 526 AATACGCTGAGCAACACCTGAAAGCTGTTGATCAAGCAATGCGACATGTTGATATCT 467  
 QY 503 TCGATCGCTGAACTTCATCATCAAGTCAAGCGCTGAAAGCGCTGACGCTTCCTCG 562  
 Db 466 TGGACAGGCTGAATTCGATCATGATCAAGTCAAGTCAAGCGTCAAGTCTTCTTG 407  
 QY 563 CTGTTGACTCTTATTCGTTGCTGCGAATAACAGATCATGACGCTGATCGGAGATCA 622  
 Db 406 CCGTGGCTCTTATTCGTTGCTGCGAATAACAGATCATGACGCTGATCGGATTA 347  
 QY 623 CCGAAGCGGCTGCTGCGCGAGCGAGTAAATCCGCGATGCTGTTAGGTCGCTG 682  
 Db 346 CAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287  
 QY 683 TGTCTGAAGCATTCGCGAGCAGCGTATCGCTGCGCGCGATCCGCTGCGAGAGA 742  
 Db 286 TGGCTGAAGCATTCGCGAGCAGCGTATCGCTGCGCGCGATCCGCTGCGAGAGA 227  
 QY 743 TCAAGTGGCTTTCATATTTGAATGCGTCCGATCCGTTGCGGAGGATCAACTTCA 802

Db	226	TGAAGTCCGGTTTGATATCTAAATTCGTACGGATCCGCTCAOGTGCATCAACTTTA	167
OY	803	TGCGCTGCCCAACCCCTTCGGGTACGAATTTGATTTGTTATCGGTAAACCGCCTGG	86
Db	166	TTGCTGCCCCAACCTCTTCACGCCCAAGATTTGATTTGGTACGGTAATATCCTTTGG	107
OY	863	AGCAACGCCCTGGAGATATCATCTACCTCGATGAGACGTTTCATTATCGCTGCGTGGA	922
Db	106	AGCACCGCCTCGAAGATATTTATCACGCCGATGAGATGTCCTATATATTTGTTGTGAGTA	47
OY	923	ATGGCCCGAGTGAGGCGCTGGTTTCTATACCTGCGCGTACCGCGGG	967
Db	46	ATGGCCCGGAGTGAACCGCAGGTTTCTACTTTTAAAGTGTGGCTGGGG	2

RESULT 9

```

Sequence, Application US/08827190
Patent No. 5858367
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medten & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-827-190-9

```

Query Match	24.0%;	Score 269;	DB 2;	Length 886;
Best Local Similarity	56.5%;	Pred. No. 4.8e-75;		
Matches 500;	Conservative	0;	Mismatches 385;	Indels 0;
			Gaps	0

QY	83	TCGCGGTACANTCATATACCAATTCGGGTGACAGACGTCGAAAGCAAGGTCATCAAA	142
Db	886	TTTGCATCCAAAGCATCAACAACAACAAACACATGATGTGAAGCAACGGTTTCGGGAAA	827
QY	143	TCAGAGCGCTGGAAACGGGTGGCGCTGATATCGTCGCTGATATCCGATACGACGATGGAGC	202
Db	826	TTTAAACGTTTGGCTGAAAGCCGGATGCCAAATCGTTGGGGTAGCATGTCCGGATGAACGGG	767
QY	203	CGCGACAGCGGTTCAAACTCATCAACACAGAGGTTAACTGCGCGCTGGTGGGTCACATCC	262
Db	766	CGGCAAAAGCCCATTCGGGATTTTAAAGAGGCATTTCATTCCTCTCGTTGGTTGACATAC	707

OY	263	ACTTGACTATTCGCACTTGGCGTGGAAAGTAGGCGAATACGGCTGATGTTGCTGGCTATTA	3222
Db	706	ATTTCGATTATTAACCTTCGCTTTCGAAGCCATTGAAGCGCGCCGCAATAAATCCGAATCA	6474
OY	323	ACCCCTGGCAATATCGGTATGAAGAGCGCTATTCGCTAGTGCGTGGTGTGACTGTGCGCGCAT	3823
Db	646	ACCCCGGCAATATCGCGCGCGGCGCAAAAAGTTGAAGGCGTGTGTTAAAGCGGCGCAAGACA	5875
OY	383	AAACATTCGCATTCGCTTTGGCGCTTAACGGCGGATCGCTGGAAAAAGATCTGCAGAAA	4424
Db	566	AAGCATTCGCATTCGAGATCGGAGTAAAGCGCGGCTCATTTGGAAAAACGATTTTAAAAA	5224
OY	443	AGTATGCGCAACCGACGCGCGGCGGTGCTGGAAATCGCCATGCGCTATGTTGATATC	5024
Db	526	AATACGGTTATCCGACTCGCATGGAAATGTAAGAAGCGCACTTCATCACATTTAAATTC	4674
OY	503	TTCGATCGCCTAACCTTCGATCAGTTCAAAATCAGCGGAAAAGCGTCAGACTCTTCTCG	5624
Db	466	TTGAGGATCTTGATTTTTCACATATTTATTTGTCACACTGAAGAGCGCTCGACGTGAACCTTG	4074
OY	563	CTGTGAGCTCTTATTCGTTGCTGCGCAAAACGATCGATCAAGCGCTTCATCTGGAGATCA	6224
Db	406	CAATCGAAGCTTATGAAGAAACAGCAGAAAGCGTTTGACTACCCGCTTCACCTCGGATCA	3474
OY	623	CCGAAGCCGGTGGTGGCGGACCGGGGAGTAAATCGCCATTTGGTTTAGTCTGCTGC	6824
Db	346	CCGAGTCAGGAACACTGTTTCCCGGCAAGTAAAGACGCGACGAGCACTCGGCGCATTTT	2874
OY	683	TGTCTGAAGGATCGGCGACACGCTGGCGGATGATGGCGGGCGGATCGCGGTGGACAGA	7424
Db	286	TAAACAAGGCAATGCGGGAACACCATGGCATTTACTTAAGCGCAGACCCCTGTAGAAAGG	2274
OY	743	TCAAGTCGTTTCGATATTTTGAATTCGCTGCTATCGCTTGGCGAGGAGTCAACTTCA	8024
Db	226	TAAAGTAGCAAGGAGAGCTTCTGAATCTTTGCGCTTAGCTCCAAATGCTGCCACGCTAA	1674
OY	803	TTCGCTGCGCGACCTGTTTCGGCTGAGGAATTTGATGTTATTCGGTAAGCGTTAAAGCGCTGG	8624
Db	166	TCTCATGCGCGCACTTGGCGCGTATTAGATGTATCTAATCAGGATTTGCCAATGAAGTGG	1074
OY	863	AGCAACGCCCTTGAAGATATCATCACTCGATGCACTTTTCGATTAAGGATGCGTGGGTGA	9224
Db	106	AAGGATATATTTCTAAGATAAAAACGCGCGATTAAGTTGCTGTCTCGCTGCGCTGTAA	474
OY	923	ATGAGCCAGGTGAGGCGCTGTTTCTACATCTCGGCGTCAACCGCG	9674
Db	46	ACGGAAGCTGGTGAAGCAGGAAGCAAGCTATATTCGAATTCGCGGCG	24

RESULT 10  
US-09-170-187-9/c

```

; Sequence 9, Application US/09170187
; Patent No. 6383745
; GENERAL INFORMATION:

```

1  
 2 APPLICANT: Rather, Philip N.  
 3 TITLE OF INVENTION: Methods For Screening For Antimicrobials  
 4 TITLE OF INVENTION: utilizing aarc And Compositions Thereof  
 5 NUMBER OF SEQUENCES: 13  
 6 CORRESPONDENCE ADDRESS:  
 7

ADDRESS: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,187  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 886 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-09-170-187-9

Query Match 24.0%; Score 269; DB 4; Length 886;  
Best Local Similarity 56.5%; Pred. No. 4,8e-75;  
Matches 500; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

QY 83 TCCTCCGTCAGTCATGACCAATACGCGTAGACAGACGTCGACCAACGTCATCAAA 142  
DB 886 TTGTCATCCAAAGCATGACACAAACACATGATGTAGAACACGCTTGGCGAAA 827  
QY 143 TCAAGGCGCTGAGAGCGGTGGCTGATGTCGTGATCCGTACCGAGACAGATGAGAC 202  
DB 826 TTAAACCGTTGGCTGAGAGCGGTGATCCCAATGCTGCGGTGATGTCGAGTGAACGG 767  
QY 203 CGGAGAAAGCGTTCAAACTCAAAACAGCAGGTTACGTCGCGTGTGCTGACATCC 262  
DB 766 CGGCAAGCGCATGCGGATATTAAGAGCGCATTTCCATTCCTCGTGTGACATAC 707  
QY 263 ACTTGACATGTCGATGCGCTGAAAGTAGCGGAATACGCGCTGATTTCTGCGATTA 322  
DB 706 ATTTGATTTAACTTGCGTTGAAAGCGCATTTGAAGCGCGCATTAATAATCCGATCA 647  
QY 323 ACCCTGGCAATTCGTTATGAAAGCGTATTCGATGTTGTCGTGCGCGGATCA 382  
DB 646 ACCCGGCAATTCGCGCGCGCGGCGGAAAGTTGAAGCGGTTGTTAAAGCGCGCAAGACA 587  
QY 383 AAAACATTCGATCCGATTTGCGTTAAAGCGCGGATCCGTGAAAAAGATCTGCAAGAAA 442  
DB 586 AAGGATTCGATCAGAAATCGGAGTAAAGCGCGGTTGATGAAAAAGCATTTAGAAA 527  
QY 443 AGTAGGCGAAGCGCGCGCGCGGCTGCTGGAATTCGCAATGCGTCAATGTTGATCATC 502  
DB 526 AATACGTTATCCGATCGCGGATGGAATGGAAGCGCATCTCATCAATTAATAATTC 467  
QY 503 TCGATCGGCTGAACTGATCGATCAAAAGTGAAGGCTGAAGCGCTGCGCTTCCTG 562  
DB 466 TTGAGGATCTGATTTTTCAGATATTAATGTCAGATGAAGGCTCTGACGTAAGACTTG 407  
QY 563 CTGTTGAGTCTTATGCTTTGCTGCAAAAACAGATGATCAGCGCTTGCATTCGGGATCA 622  
DB 406 CAATGAGGCTTATGAAAAAGCAGGAAAGCGTTGACTACCGCTTCACCTCGGATCA 347  
QY 623 CGAAGCGCGGTGGCGCGCGCGGCGGAGTAAATCCGCCATTTGGTTAGTCTGCTGC 682  
DB 346 CCGATTCAGGAACACTGTTTGGCGCACAGTAAAGAGCGCACAGACTCGCGCATTT 287  
QY 683 TGTCTGAGGCGATCGGCGGACAGCTGCGGTGATCGTGGCGCGCGATCCGCGCAAGANA 742  
DB 286 TAAGCAAGGCAATCGGAGACACCATGCGATTTCACTAAAGCCCTGTAGGAAGAG 227  
QY 743 TCAAGTCGCTTATGATATTTTGAATGCTGCTGATCCGTTGCGGAGGATCAACTTCA 802  
DB 226 TAAAGTAGCAAGGAGGCTTCTGAATCTTTGGGCTAGACCCCAATGCTGCACAGCTCA 167  
QY 803 TCGCTGCCCGACCTGTTGCGCTCAGGAATTTGATGTTATCGGTAAGCGGCTGG 862

DB 166 TCTCATGCCCGGCTTGGCGCGCGCTATGATGATCTATACGATTTGCCATGAGTGG 107  
QY 863 AGCAAGCGCTGAGATATATCATCTACCTCCGATGAGAGCTTTCATTCGCGCTGCGTGA 922  
DB 106 AAGATATATTTCTAAGATTAAGAGCGCGCATTAAGATTGCTTTCTCGGCTCGCTGTA 47  
QY 923 ATGGCCAGGTGAGCGCGCTGTTCTACACTCGGCGTCAACCGCGC 967  
DB 46 ACGACCTGTGTACAGCAGAGAAAGCTGATATCGAATGCGCGCGC 2

RESULT 11  
US-09-221-017B-309  
Sequence 309, Application US/09221017B  
Patent No. 644799  
GENERAL INFORMATION:  
APPLICANT: ROSS, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: MONROY, Gladys H.  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 309:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...1083  
US-09-221-017B-309

Query Match 5.0%; Score 55.4; DB 4; Length 1083;







[illegible]

Search completed: May 10, 2003 18:49:30  
Job time : 7711 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 16:41:08 ; Search time 197 Seconds  
(Without alignments)  
7058.816 Million cell updates/sec

Title: US-09-921-992-3  
Perfect score: 1119  
Sequence: 1 atgcataaccaggtcccaat.....ttcagcaggttcgaataaa 1119

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352456 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	10 US-09-921-992-3	Sequence 3, Appl1
2	403.6	36.1	640681	10 US-09-921-992-3	Sequence 1, Appl1
3	293.4	26.2	1083	10 US-09-921-992-3	Sequence 1692, Ap
4	243.8	21.8	1164	9 US-09-921-992-3	Sequence 110, Ap
5	239.4	21.4	1134	9 US-09-921-992-3	Sequence 2205, Ap
6	211.8	18.9	1180	10 US-09-921-992-3	Sequence 195, Ap
7	70.6	6.3	384	10 US-09-921-992-3	Sequence 21, Appl
8	68.4	6.1	566	10 US-09-921-992-3	Sequence 23, Appl
9	56.8	5.1	670	10 US-09-921-992-3	Sequence 22, Appl
10	56.8	5.1	348	10 US-09-921-992-3	Sequence 22, Appl
11	52.6	4.7	293	10 US-09-921-992-3	Sequence 6102, Ap
12	52.6	4.7	456	10 US-09-921-992-3	Sequence 25, Appl
13	50	4.5	403	10 US-09-921-992-3	Sequence 26, Appl
14	48.8	4.4	564	10 US-09-921-992-3	Sequence 24, Appl
15	48.4	4.3	379	10 US-09-921-992-3	Sequence 16, Appl
16	48.4	4.3	528	10 US-09-921-992-3	Sequence 15, Appl
17	48.4	4.3	2520	10 US-09-921-992-3	Sequence 1, Appl1
18	48.2	4.3	353	10 US-09-921-992-3	Sequence 39, Appl
19	48.2	4.3	464	10 US-09-921-992-3	Sequence 33, Appl

20	47.2	4.2	601	10 US-09-921-992-3	Sequence 11, Appl
21	45.4	4.1	613	10 US-09-921-992-3	Sequence 32, Appl
22	45.2	4.0	432	10 US-09-921-992-3	Sequence 14, Appl
23	45	4.0	938	10 US-09-921-992-3	Sequence 13, Appl
24	43.6	3.9	443	10 US-09-921-992-3	Sequence 12, Appl
25	43.4	3.9	291	10 US-09-921-992-3	Sequence 40, Appl
26	41.8	3.7	577	10 US-09-921-992-3	Sequence 43, Appl
27	41	3.7	422	10 US-09-921-992-3	Sequence 28, Appl
28	41	3.7	430	10 US-09-921-992-3	Sequence 29, Appl
29	40.2	3.6	3675	10 US-09-921-992-3	Sequence 2, Appl1
30	39.6	3.5	671	9 US-10-184-634-346	Sequence 346, App
31	39.6	3.5	671	9 US-10-184-634-346	Sequence 346, App
32	38.8	3.5	388	10 US-09-921-992-3	Sequence 44, Appl
33	37.2	3.3	511	10 US-09-921-992-3	Sequence 36, Appl
34	37.2	3.3	705	10 US-09-921-992-3	Sequence 34, Appl
35	36.6	3.3	395	10 US-09-921-992-3	Sequence 18, Appl
36	35.6	3.2	211	10 US-09-921-992-3	Sequence 45, Appl
37	35	3.1	1383	9 US-09-970-989-6	Sequence 6, Appl1
38	35	3.1	1383	10 US-09-967-768A-189	Sequence 189, App
39	35	3.1	1522	9 US-09-954-531-612	Sequence 612, App
40	35	3.1	1522	10 US-09-962-436-20	Sequence 20, Appl
41	34.8	3.1	3624	9 US-10-108-605-216	Sequence 216, App
42	34.2	3.1	1223	10 US-09-812-102-9	Sequence 9, Appl1
43	33.6	3.0	425	9 US-10-025-380-589	Sequence 589, App
44	33.6	3.0	425	10 US-09-922-217-589	Sequence 589, App
45	33.6	3.0	425	10 US-09-833-263-589	Sequence 589, App

## ALIGNMENTS

RESULT 1  
US-09-921-992-3  
Sequence 3, Application US/09921992  
Patent No. US2002069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronol, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohmer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
APPLICANT: Venkatesh, Mylavarapu  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 3  
LENGTH: 1119  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1119)  
US-09-921-992-3  
Query Match 100.0%; Score 1119; DB 10; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACCGGCTCCATTCAGTGAAGAAATCAACAGTATTTACGTTGGAGATGTG 60  
|||||  
Db 1 ATGCATACCGGCTCCATTCAGTGAAGAAATCAACAGTATTTACGTTGGAGATGTG 60  
|||||  
QY 61 CCGATTGGCGATGGTGGTCCCATGCGGTACAGTCAATGACCAATPAGCGGTACGAC 120  
|||||  
Db 61 CCGATTGGCGATGGTGGTCCCATGCGGTACAGTCAATGACCAATPAGCGGTACGAC 120  
|||||  
QY 121 GTCGAACCAAGGTCAATCAATCAATCAAGGCGCTGGAGCGCTTGGCGGTGATATCGTCCGT 180  
|||||



QY 799 TTCATCGCCTGCCGACCTGTCGCGTCAAGAAATTGATGTTATGCGTACGGTTAACGG 858  
 Db 315073 TTTATTTGCTTGTCTACTGTTCTTAGACAAAGAAATTGATGTTATGATACGTAATCA 315132  
 QY 859 CTGGAGCAACGCGCTGGAGATATCATCACTCCGATGACGTTTCATTAATCGCTGCGT 918  
 Db 315133 CTAGAAAAAATCTAGAAAGATATCTGACCTCCATAGATGATCAATTAATGTTGCGT 315192  
 QY 919 GTGATGCGCCGAGTGAGCGCGTGTCTCTACACTCGCGGTCACCGCGGCAACAAGAA 978  
 Db 315193 GTTAATGGAATATGATGATCTTAATAATACCACTTTAGTCTAGCAGGAATCATAGAAA 315252  
 QY 979 AGCGGCTCTATAGATGCGGT---GGCAAGACCGCTGAGACAGCATATGATC 1035  
 Db 315253 AGTCATTTTATGAGAGCGAGGTAGACAAAAGAAAAAATTAAGAGAAATTAATA 315312  
 QY 1036 GACCAAGCTGGAAGACGCGATTCGTCGAAA 1065  
 Db 315313 GAAAAAATGAAATTAATAATTCGAAAAAA 315342

RESULT 3

US-09-974-300-1692  
 : Sequence 1692, Application US/09974300  
 : Patent No. US20020146721A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Berka, Randy M.  
 : APPLICANT: Clausen, Ib Groth  
 : TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 : FILE REFERENCE: 10085.500-US  
 : CURRENT FILING DATE: 2001-10-05  
 : PRIOR APPLICATION NUMBER: 09/680,598  
 : PRIOR FILING DATE: 2000-10-06  
 : PRIOR APPLICATION NUMBER: 60/279,526  
 : PRIOR FILING DATE: 2001-03-27  
 : NUMBER OF SEQ ID NOS: 8481  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 1692  
 : LENGTH: 1083  
 : TYPE: DNA  
 : ORGANISM: Bacillus licheniformis  
 : US-09-974-300-1692

Query Match 26.2% Score 293.4; DB 10; Length 1083;  
 Best Local Similarity 56.6% Pred. No. 1.2e-90;  
 Matches 543; Conservative 0; Mismatches 416; Indels 0; Gaps 0;  
 QY 22 CAACGTAGAAATCAACAGATATTACGTGGGATGTCGCGATGGCGATGTCGCC 81  
 Db 21 CATGCTACAAAAGCCGTCGCTTAAGTGGACCTTTAACAATAGCGGCAATACGA 80  
 QY 82 ATCGCGTACAGTCCATACCAATACGGGTACGACAGCTGCAAGACGGTCAATCA 141  
 Db 81 GTCGTCATTTCAAGCATCAACAACAACACATGACGTGAGCAACCGTCCGCA 140  
 QY 142 ATCAAGGGCGTGAAGCGGTGGCGCTGATATCGTCCGTATCGTACCGCATGAGC 201  
 Db 141 ATCAACAGACTGCGGAGACAGGATGTCAAAATGTCGCGCTGCTGCTGATGAACGG 200  
 QY 202 GCGGAGAGCGTTCAAATCTCAACAGCAGGTTAAACGTCGCGCTGGTGGCTGACATC 261  
 Db 201 GCTGCGAGCGCATTCGAGAGATCAAAAAGGAGATATCATCCCTGTGTGCTGATAT 260  
 QY 262 CACTTCGACTATCGATTCGCTGAAAGTAGCGGAATACGGCGTGGATTTGCTGCTAT 321  
 Db 261 CATTTCAACTATAATTTGCGATTTAAAGCATGCAAGGCGGCGCATTAATAATCCGATC 320  
 QY 322 AACCTGCGCAATATGCTGATATGCAAGCGGTATTCGATGTTGGTGTGCTGCGCGAT 381  
 Db 321 AATCGGGTAACTGCGCGCGCGGCAAAAGTTGAAGCGGTGCTCAACGCAAGGA 380

QY 382 AAAACATTCGATCCGATATTTGCGTTAAACGCGGATCGCTGGAAAAAGATCTGCAAGA 441  
 Db 381 AAGGCGATTTCCGATCCGATGCGGCTCAATGCAAGGCTCTCTGAAAAAAGCAATCCTTGAG 440  
 QY 442 AAGTATGCGAAGCAGCGCGCGAGCGTGTGTAATCTCCATCGCTCATCTTGATCAT 501  
 Db 441 AAGTACGGCTATCCAGACAGCAGACGCGATGTCGAAGCGCCCTCACACATTAATAAT 500  
 QY 502 CTCGATCGCTGAACCTTGATGATTCATCAAGTCACGCGTAAGAGCGCTGACGCTTCC 561  
 Db 501 CTGAGAGATCTCGATTTCCACGATATCATCTGACACATGAAGCGCTCTGATGTAACCTG 560  
 QY 562 GCTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCATACAGCTTGATCTGAGGATC 621  
 Db 561 GCGATTTAGGCGATATGAANAAGCGCTAAAGCTTTGATATTCGCTTATTTAGCATC 620  
 QY 622 ACCGAAGCGGTGTGTCGCGGACGCGGCGAGTAAATCCGCAATGCTTTAGTCTGTG 681  
 Db 621 ACCGAATCGGAGACGCTGTTGCGGTACAGTGAAGAGCGCGGCTCTCGCGCGATC 680  
 QY 682 CTGCTGAAGCGATGCGGACACGCTGCGCGTATGCGCGCGGATCCGCTGCAAGAG 741  
 Db 681 CTTTCAAAAGGATCGCAATATCATTTGCGGATTTCTTAAGCGCGGACCGCTCGAAGA 740  
 QY 742 ATCAAGTCTGTTGATATTTTGAATCGCTGCTATCCGTCGAGAGATCAACTTC 801  
 Db 741 GTAAAGTCTGCCAGGAGCTCTGAAATCTTTGCGGCTGCTCCAAATGCGCAACATTTG 800  
 QY 802 ATGCGCTGCCGACCTGTTGCGGTGAGCAATTTGATGTTATGCTGATGCTTAAACGCTG 861  
 Db 801 ATTCTCTCCCGCATTTGCGGCGGATCGAAATCGATTTGATGCTGATGCTCAATGATTC 860  
 QY 862 GAGCAAGCGCTGAGATATCATCTCGGAGCAAGCTTGCATATGCGGCTGCTG 921  
 Db 861 GAGATTCATCGCAATATCAAGCGCGGATCAAGTGAAGTCTGCTGCTGCTGCTGCTC 920  
 QY 922 AATGCCAGATGAGCGCGTGTCTTCTACACTCGGCGTACCGCGGCAACAAGAAAG 980  
 Db 921 AAGGTCGCGGAGAAAGCGGCGGCAAGCCGATATTCGATGCGCGGCGGCAAGG 979

RESULT 4

US-09-712-363-110  
 : Sequence 110, Application US/09712363  
 : Patent No. US20020164588A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Eisenberg, David  
 : APPLICANT: Rotstein, Sergio H.  
 : APPLICANT: Marcotte, Edward M.  
 : TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 : FILE REFERENCE: 07419-032001  
 : CURRENT FILING DATE: US/09/712,363  
 : PRIOR APPLICATION NUMBER: 2000-11-13  
 : PRIOR FILING DATE: PCT/US00/02246  
 : PRIOR APPLICATION NUMBER: 60/01-28  
 : PRIOR FILING DATE: 2000-01-28  
 : PRIOR APPLICATION NUMBER: 60/179,531  
 : PRIOR FILING DATE: 2000-02-01  
 : PRIOR APPLICATION NUMBER: 60/117,844  
 : PRIOR FILING DATE: 1999-01-29  
 : PRIOR APPLICATION NUMBER: 60/118,206  
 : PRIOR FILING DATE: 1999-02-01  
 : PRIOR APPLICATION NUMBER: 60/126,593  
 : PRIOR FILING DATE: 1999-03-26  
 : PRIOR APPLICATION NUMBER: 60/134,093  
 : PRIOR FILING DATE: 1999-05-14  
 : PRIOR APPLICATION NUMBER: 60/134,092  
 : PRIOR FILING DATE: 1999-05-14  
 : PRIOR APPLICATION NUMBER: 60/165,124  
 : PRIOR FILING DATE: 1999-11-12  
 : PRIOR APPLICATION NUMBER: 60/165,086  
 : PRIOR FILING DATE: 1999-11-12





QY 436 CAAGAAAGTATGCGACCGACGCCGCTGCTGGAATCTGCCATGCGTCACTTT 495  
DB 430 GCAAAATACACGCAAAACCCAGAGAGCTCTGCTGGAATCCGCAATGCGGAGCC 489  
QY 496 GATCATCTGCGATCGCGTGAACCTTCATGCTCAAGATGAGGTGAACCGTCTGCGTC 555  
DB 490 GCGCTGTTTGAAGACACGCGCTTCGCGCATGCGCATCTTCTGTGAAGCATCCGACCA 549  
QY 556 TTCTCGCTGTTGAGTCTTATCTGTTGCGCAAAACAGATGACAGCGCTTGATCTG 615  
DB 550 GTACTCATGCTGAGAGCGCTACCGCAGCTCGGTGAACAAAGGACATCCACTGACCTTC 609  
QY 616 GGGATACCGAAGCGGTGCTGCGCAGCGGGGCACTAAATCCGCCATTGTTAGST 675  
DB 610 GGTGTACTGAAGCTGTGCTCCAAAGTTTCATGAGGAAATCAAGTCTTCCCTACATTCGCG 669  
QY 676 CTGCTGCTGTGTAAGGCGTGGCGACAGCTGCGGTATGCTGCGCGCGCATCCGCTG 735  
DB 670 GCTCTGCTGTCCAGAGGCTGCGCGACACTATCCGTGTCTCTTCTGCTGACCCAGT 729  
QY 736 GAGAGATCAAAAGTGGTTTCATATTTGAATGCGTGCCTATCCGTTCCGAGAGATC 795  
DB 730 GAAGAATCAAGGTGGGAGCAGATCTGCAAGTCCCTCAACCTGGCCAGCAAGCTG 789  
QY 796 AACTTCATGCGCTGCGCGACCTGTTGCGCTCAGGAATTTGATGTTACGCTTAC 855  
DB 790 GAATTCGTCTCTGCTCCATCATGTGCGCGCAGAGTGTGATCTACCTTGTGTA 849  
QY 856 GCGGTGAGACAGCGCTGGAATATCATCTCGATGAGCGTTTGCATTTACGCTG 915  
DB 850 GAGTACCGAAGCATGACGCGCATGGAATTCACCTGCGCTTCATGGTTGC 909  
QY 916 GTGTGATGAGCGGAGGTGAGCGCTGTTTCTACACTGCGGCTCAACCGCGCAAG 975  
DB 910 GTTGTAAAGCGCCAGGTGAGCGCTGCGCATGAGCTGCGTGTGATCCGTAACGCGC 969  
QY 976 AAAAGCGCGCTCTATGAAGATGGCG 1000  
DB 970 AAGGCGCAGATCTTGTCAAGGCG 994

RESULT 6  
US-09-881-752A-195  
; Sequence 195, Application US/09881752A  
; Patent No. US20020115078A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleantous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Oomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the  
; FILE REFERENCE: 06132/041002  
; CURRENT APPLICATION NUMBER: US/09/881,752A  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 08/833,457  
; PRIOR FILING DATE: 1997-04-01  
; NUMBER OF SEQ ID NOS: 370  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 195  
; LENGTH: 1180  
; TYPE: DNA  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (51)...(1127)  
US-09-881-752A-195

Query Match 18.9%; Score 211.8; DB 10; Length 1180;  
Best Local Similarity 52.0%; Pred No. 2e-62;  
Matches 524; Conservative 0; Mismatches 477; Indels 6; Gaps 2;

QY 31 AATCAACAGATATTTACGTTGGGAATGTCCGATTTGGCGATGTCTCCATCCGCTGA 90  
DB 69 AAGACCAAGCAATTTTATGCTGTGGCGCATATGAGGGGTGATGTCTCCATATAGCAG 128  
QY 91 CAGTCATGACCAATATCCGCTACGACAGCTGGAAGCAAGCGTCAATCAATCAAGCG 150  
DB 129 CAAGCATGACCTTTAGCAAAAACCGCTGATTTGAAGCACTAAATCAATTTGACAGA 188  
QY 151 CTGAAGCGGTTGGGCGTATGCGTGTATCGTACGACATGAGAGCGCGACAGA 210  
DB 189 CTCAAACTGCGCGGCGCATTTAGTGAAGGCGGGAGTAAAGAAAAGACGCTCTA 248  
QY 211 GCGTTCAAACTCATCAACAGCAGGTTAACGTGCGCGCTGCTGATCCATCCATCGAC 270  
DB 249 GCGTTAAAGAAATGAAAAAAGTGTCCCTTGTGCTTATGCTGATATTCATTTCCAT 308  
QY 271 TATCGCATTTGCGCTGAAGTACCGGATATGCGCTGATGTTCTGCTATTAACCTGCG 330  
DB 309 TATAAATTCGCTCTCATGCGCTCAAA--CGGTGATGCGATCAGGATTAACCCCGGA 365  
QY 331 AATATCGTAAATGAAGAGCGTATTCGATGTTGACTGTGACGTGCGCGCATTAACAT 390  
DB 366 AACATCGGCTTAAGAGAGATCAAGCGGTGTTGATGCTTTGAAGAAAAAACAAT 425  
QY 391 CCGATCCGTTATGGCGTTAACCGCGGATCGTGAAAAAGATCTGCAGAAAGATATGCG 450  
DB 426 CCTATAAGAAATTTGGCGTGAATGCTGGAGATTTAGAAAACGATTTGATCAAAAATA--C 482  
QY 451 GAACGAGCGCGCAGCGCTGCTGGAATCTGCAATGCGTCAATGTTGATCATCTCGATCG 510  
DB 483 GAGCCCAACCCAAAAGGCAATGTAAGAACGCTTTGTATACGCCAAATTTTAAAGAT 542  
QY 511 CTGAATTCGATCACTTCAAGTCAAGCGTGAAGGCTGACGCTTCTCTGCTGTTAG 570  
DB 543 TTGATTTTACCAATTTTAAATTTCTTTAAAGAGAGGATGTATTCGACCAATAGA 602  
QY 571 TCTTATCGTTTCTGCGCAAAACAGATCGATCGATCGCTGATCTGGGATCACCGAGCG 630  
DB 603 GCTTACAGAGATGCTTCCGCTCTGTGATCTATCTTCCATTTGGGGGTACGAGCGCG 662  
QY 631 GGTGGCGCGGAGCGGCGCATTAATCCGCTATGTTAGTCTGTGCTGTCTGA 690  
DB 663 GGGAACTTTTATGCTCCAGTATCAATCCGCTATGCTTATGAGGGGCTTTTAAATGAG 722  
QY 691 GGCATCGCGACACCGCTGCGTATGCTGCGCGCGATCCGCTGGAAGAGATCAAGTC 750  
DB 723 GGCATTTGGGATACGATGCGCTATTCATCAGAGGGAATTTGAATAATCAAAAGTG 782  
QY 751 GGTTCGATATTTGAATGCGTGTATCGCTGCGAGGAGATCACTTCATCGCTGCG 810  
DB 783 GCCAGAGCAATTTTACGCGATAGCGGCGGTTGAAGAAAGGAGATTAATGATTTCTG 842  
QY 811 CGGACTGTGCGCGTACAGAAATTTGATGATTCGATGAGCGTTAACCGCTGAGACAGCG 870  
DB 843 CCGACTGTGCGCGATGGAACCAATTTAGTGAATGCGCATCAAGGTAAAAAACCG 902  
QY 871 CTGAAGATATCATCTACCTGATGAGCGTTTATGATGCGTGTGATGAGTGAAGCGCA 930  
DB 903 TTAAGCCACATCAAAACCTTTAGACATTTACGATGATGAGTGTGCTGTGATCTTTG 962  
QY 931 GGTGAGCGCGTGTCTTACACTGCGCTCACCGCGCGACAAAGAAAGGCGCTCTAT 990  
DB 963 GGTGAAGCCAGCATGACATGCGATGCGCTTTTGGGAATCGGAGCGTTTGTATCAT 1022  
QY 991 GAAGATGCGTGGCAAGAGCGCTGAGCAACACGATATGATGA 1037  
DB 1023 AAAGAGGTTAAATTCATTCACAACTGCGTGAAGGATTTATTTGA 1069

RESULT 7  
US-09-921-992-21/C  
; Sequence 21, Application US/09921992



Db 403 GGTTCGGGAAACCATGTGTGAATATGAGCGCTCTCTCTCTGCGGCG 452

RESULT 10  
US-09-974-300-6102

Sequence 6102, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085.500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 6102  
LENGTH: 348  
TYPE: DNA  
ORGANISM: Bacillus clausii  
US-09-974-300-6102

Query Match  
Best Local Similarity 54.1%; Score 56.8; DB 10; Length 348;  
Matches 112; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 764 TGAATCGTGGTATCCCTGCGGAGGATCACTTATCGCTGCCCGACCTGTTGC 823  
Db 1 TGAATCGTGGTATCCCTGCGGAGGATCACTTATCGCTGCCCGACCTGTTGC 60  
QY 824 GTGAGGAATTTGATTTATCGGTACGCGTGAACCGCTGGAAGATATCA 883  
Db 61 GGTATAGATTTGATTTATCGGTACGCGTGAACCGCTGGAAGATATCA 120  
QY 884 TCACCTCCGATGACGTTTGCATTTATCGCTGCTGTGATGGCCCAAGTGAAGCTCG 943  
Db 121 AAGCACCAGTAAAGAGCGCTGCTGTGATGGCCCAAGTGAAGCTCG 180  
QY 944 TTTCTACACTCGGCGTACCGCGC 967  
Db 181 AAGCAGATATCGGTATTCGCGAG 204

RESULT 11

US-09-921-992-25  
Sequence 25, Application US/09921992

Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronati, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohrer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
APPLICANT: Venkatesh, Mylavaram;  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 25  
LENGTH: 293  
TYPE: DNA  
ORGANISM: Zea mays

US-09-921-992-25

Query Match  
Best Local Similarity 4.7%; Score 52.6; DB 10; Length 293;  
Matches 109; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 25 CGTAGAAATCAACACGATTTTACGTTGGAAATGCGGATGGCATGCTCCATC 84  
Db 10 CGAGGAAATCAACACGATTTTACGTTGGAAATGCGGATGGCATGCTCCATC 69  
QY 85 GCCGACAGTCCATGACCAATACGCGTACGAGCGTGAAGCAAGGTCATCAATC 144  
Db 70 AGGATTCAAACATGACGATTCAGATTCAGAGTGTTCGAAAACGTTAGAGAGTG 129  
QY 145 AAGCGCTGGAACGCGTGTGATTCGCTGTATTCGTTACCGACGATGACCGG 204  
Db 130 ATGAGATGACGATTAAGGAGCTGATCTTTAGAAATACAGTCCAGGTAAGAGAA 189  
QY 205 GCAGAGCGTTCAACATCATCA 227  
Db 190 GCTGATGCTGCTTTGAGATCAA 212

RESULT 12

US-09-921-992-26  
Sequence 26, Application US/09921992

Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronati, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohrer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
APPLICANT: Venkatesh, Mylavaram;  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 26  
LENGTH: 456  
TYPE: DNA  
ORGANISM: Zea mays  
US-09-921-992-26

Query Match  
Best Local Similarity 4.7%; Score 52.6; DB 10; Length 456;  
Matches 109; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 25 CGTAGAAATCAACACGATTTTACGTTGGAAATGCGGATGGCATGCTCCATC 84  
Db 10 CGAGGAAATCAACACGATTTTACGTTGGAAATGCGGATGGCATGCTCCATC 69  
QY 85 GCCGACAGTCCATGACCAATACGCGTACGAGCGTGAAGCAAGGTCATCAATC 144  
Db 70 AGGATTCAAACATGACGATTCAGATTCAGAGTGTTCGAAAACGTTAGAGAGTG 129  
QY 145 AAGCGCTGGAACGCGTGTGATTCGCTGTATTCGTTACCGACGATGACCGG 204  
Db 130 ATGAGATGACGATTAAGGAGCTGATCTTTAGAAATACAGTCCAGGTAAGAGAA 189  
QY 205 GCAGAGCGTTCAACATCATCA 227  
Db 190 GCTGATGCTGCTTTGAGATCAA 212

RESULT 13

US-09-921-992-24  
Sequence 24, Application US/09921992

Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronat, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohmer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
PRIORITY FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 24  
LENGTH: 403  
TYPE: DNA  
ORGANISM: Zea mays  
US-09-921-992-24

Query Match 4.5%; Score 50; DB 10; Length 403;  
Best Local Similarity 53.0%; Pred. No. 1e-06;  
Matches 107; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 26 GTAGAAATCAACAGTATTTACGTTGGAGATGCGGATGCGGATGCTCCCATCG 85  
DB 11 GGAGGAAACCTGAACTGTGATGTGGGATGTGCCACTTGGAGATGATCCCAATA 70  
QY 86 CCGTACAGTCCATGACCAATACGCGTACGACAGAGCTGGAAGCAAGGTCATCAATCA 145  
DB 71 GGATTCAAACCATGACAGCTGACATGACCAAGATGTTGCAAAACAGTAGAGAGTGA 130  
QY 146 AGGCGCTGGAAGCGGCTTGATATCGCGGTATCCGTATCCGTCCACCATGAGCGCG 205  
DB 131 TGAGATGACGAGATTAAGGAGGCTGATCTTTGTAATTAACAGTCCAGGATGGAAGAG 190  
QY 206 CAGAAGCGTTCAAACTCATGATA 227  
DB 191 CTGATGCTGCTTGATGATGATA 212

RESULT 14  
US-09-921-992-35  
Sequence 35, Application US/09921992  
Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronat, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohmer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
PRIORITY FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 35  
LENGTH: 564  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1..564)  
OTHER INFORMATION: unsure at all n locations  
US-09-921-992-35

Query Match 4.4%; Score 48.8; DB 10; Length 564;  
Best Local Similarity 51.8%; Pred. No. 3.2e-06;  
Matches 172; Conservative 0; Mismatches 142; Indels 18; Gaps 2;

QY 26 GTAGAAATCAACAGTATTTACGTTGGAGATGCGGATGCGGATGCTCCCATCG 85  
DB 131 GGAGGAAACCAACAGAGTGTGATGTAAGTGGCTATTTGAGGAGACATCTTA 190  
QY 86 CCGTACAGTCCATGACCAATACGCGTACGACAGAGCTGGAAGCAAGGTCATCAATCA 145  
DB 191 GAATTCAGACCATGACTACGACCTGACATAGATGTTGCTGGGACAGTTGAACAGTGA 250  
QY 146 AGGCGCTGGAAGCGGCTTGATATCGCGGTATCCGTATCCGTCCACCATGAGCGCG 205  
DB 251 TGAGATGACGAGATTAAGGAGGCTGATGATGAGGATTAACAGTCAAGGAAAGAG 310  
QY 206 CAGAAGCGTTCAAACTCATGATAACAGCAGGTT-----AACGTCGCGCTGG 250  
DB 311 CTGATGCTGCTTTGATGATTAATAAACCCCTGTTCAAGAAATTACACATCTCGTGG 370  
QY 251 TGGCTGACATCCACTTGACTATCGCATTTGCCCTGAAAGTAGCGGATACGCGCTGATT 310  
DB 371 TGGCTGATATTCATTTGCTCCCTGCTGCTGCTGCTGCGGATGAGTATTTGCTGATA 427  
QY 311 GTCGCTATTATTAACCTGCAATATCGGTAAT 342  
DB 428 AGATTCGTGTAACCCGGAATTTTGTCTGAT 459

RESULT 15  
US-09-921-992-16  
Sequence 16, Application US/09921992  
Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronat, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohmer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentin, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
PRIORITY FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/223,483  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 16  
LENGTH: 379  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-921-992-16

Query Match 4.3%; Score 48.4; DB 10; Length 379;  
Best Local Similarity 58.2%; Pred. No. 3.5e-06;  
Matches 85; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 598 GATCAGCCGTTGATCTGCGGATCACCGAAGCGGTGCGCGGACGCGGACAGTA 657  
DB 215 GATTATCTTTGATGATTTGGAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 274  
QY 658 TCCGCAATGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717  
DB 275 TCTGCAATGGAATTTGGAGCGCTTTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 334  
QY 718 CTGGGCGCGATCCGCTCGAAGAGAT 743  
DB 335 CTGACGAGGACCAACAGAGAGAGAT 360

Mon May 12 08:44:42 2003

Search completed: May 10, 2003, 17:56:02  
Job time : 830 secs

us-09-921-992-3.rnpb

